Average performance of the sparsest approximation using a general dictionary

François Malgouyres⋆ and Mila Nikolova⋄

⋆ LAGA/L2TI, Université Paris 13, CNRS, 99 avenue J.B. Clément, 93430 Villetaneuse, France;
(33/0) 1-49-40-35-83, malgouy@math.univ-paris13.fr

⋄ CMLA, ENS Cachan, CNRS, PRES UniverSud, 61 Av. President Wilson, F-94230 Cachan, France
(33/0) 1 47 50 59 08 nikolova@cmla.ens-cachan.fr

March 4, 2008

Key words: compression; approximation ; best K-term approximation; constrained minimization; dictionary; ℓ_0 norm; estimation; frames; measure theory; nonconvex functions; sparse representations.

AMS class: 41A25, 41A29, 41A45, 41A50, 41A63.

Abstract

We consider the minimization of the number of non-zero coefficients (the ℓ_0 "norm") of the representation of a data set in terms of a dictionary under a fidelity constraint. (Both the dictionary and the norm defining the constraint are arbitrary.) This (nonconvex) optimization problem naturally leads to the sparsest representations, compared with other functionals instead of the ℓ_0 "norm".

Our goal is to measure the sets of data yielding a K-sparse solution—i.e. involving K non-zero components. Data are assumed uniformly distributed on a domain defined by any norm—to be chosen by the user. A precise description of these sets of data is given and relevant bounds on the Lebesgue measure of these sets are derived. They naturally lead to bound the probability of getting a K-sparse solution. We also express the expectation of the number of non-zero components. We further specify these results in the case of the Euclidean norm, the dictionary being arbitrary.

1 Introduction

1.1 The problem under consideration

Our goal is to represent observed data $d \in \mathbb{R}^N$ in a economical way using a dictionary $(\psi_i)_{i\in I}$ on $\mathbb{R}^N$, where I is a finite set of indexes and

$$\operatorname{span}\left\{\psi_i : i \in I\right\} = \mathbb{R}^N. \tag{1}$$

We study the sparsest representation where the (unknown) coefficients $(\lambda_i)_{i\in I}$ are estimated by solving the constraint optimization problem $(\mathcal{P}_d)$ given below:

$$(\mathcal{P}_d) : \quad \left\{ \begin{array}{l} \text{minimize}_{(\lambda_i)_{i\in I}} \ell_0\big((\lambda_i)_{i\in I}\big), \\ \text{under the constraint : } \quad \left\| \sum_{i\in I} \lambda_i \psi_i - d \right\| \le \tau, \end{array} \right. \tag{2}$$

with

$$\ell_0((\lambda_i)_{i\in I}) \stackrel{\text{def}}{=} \#\big\{i \in I : \lambda_i \neq 0\big\},$$

where $\#$ stands for cardinality, $\|.\|$ is an arbitrary norm and $\tau > 0$ is a fixed parameter. Let us emphasize that for any $d \in \mathbb{R}^N$, the constraint in $(\mathcal{P}_d)$ is nonempty thanks to (1) and that the minimum is reached since ℓ_0 takes its values in the finite set $\{0, 1, \ldots, \#I\}$.

Given the data d, the norm $\|.\|$, the parameter τ and the dictionary, the solution of $(\mathcal{P}_d)$ is the sparsest possible, since the objective function ℓ_0 in (2) minimizes the number of all non-zero coefficients in the set $(\lambda_i)_{i\in I}$ without penalizing them.

The function ℓ_0 is sometimes abusively called the ℓ_0-norm. It can equivalently be written as

$$\sum_{i\in I} \varphi(\lambda_i) \quad \text{where} \quad \varphi(t) = \left\{ \begin{array}{lll} 0 & \text{if} & t = 0 \\ 1 & \text{if} & t \neq 0 \end{array} \right. \tag{3}$$

The function φ is discontinuous at zero and $\mathcal{C}^\infty$ beyond the origin, and has a long history. It was used in the context of Markov random fields by Geman and Geman 1984, cf. [8] and Besag 1986 [1] as a prior in MAP energies to restore labeled images (i.e. each λ_i belonging to a finite set of values):

$$\mathcal{E}(\lambda) = \Big\| \sum_{i\in I} \lambda_i \psi_i - d \Big\|_2^2 + \beta \sum_{i\sim j} \varphi(\lambda_i - \lambda_j), \tag{4}$$

where the last term in (4) counts the number of all pairs of dissimilar neighbors i and j, and $\beta > 0$ is a parameter. This label-designed form is known as the Potts prior model, or as the multi-level logistic model [2, 11]. Guided by the *Minimum description length* principle of Rissanen, Y. Leclerc proposed in 1989 in [10] the same prior to restore piecewise constant, real-valued images. The hard-thresholding method to restore noisy wavelet coefficients, proposed by Donoho and Johnstone in 1992, see [6], amounts to minimize for each coefficient λ_i a function of the form $\|\lambda_i - g_i\|_2^2 + \beta\varphi(\lambda_i)$ where the noisy coefficients read $g_i = \langle \psi_i^*, d\rangle$, $\forall i \in I$ where $(\psi_i)_{i\in I}$ is a wavelet basis. Very recently, the energy (4) was successfully used to reconstruct 3D tomographic images by using stochastic continuation by Robini and Magnin [19]. Let us notice that even though the problem $(\mathcal{P}_d)$ in (2) and the minimization of $\mathcal{E}$ in (4) are closely related, there is no rigorous equivalence in general.

The context of digital image compression is of a particular interest, since it is typically the problem we are modeling in the paper. In compression, one considers different classes of images. Those digital images live in $\mathbb{R}^N$ and are obtained by sampling an analogue image. Their distribution in $\mathbb{R}^N$ is one of the main unknown in image processing and, in practice, we only know some realizations of this distribution (i.e. some images). Given this (unknown) distribution, the goal of image compression is to build a coder (that encodes elements of $\mathbb{R}^N$) which assigns a small code to images. Typically, we want for every image $d \in \mathbb{R}^N$

$$\mathbb{P}\left(length(code(d)) = K\right)$$

to be as large as possible for K small, and small for K large. We also want the decoder to satisfy $decode(code(d)) \sim d$.

The link with the problem $(\mathcal{P}_d)$, in (2), is that the current image compression standards (JPEG, JPEG2000) encode quantized versions of the coordinates of the image in a given basis. Moreover, most of the gain is made by choosing a basis such that the number of non-zero coordinates (after the quantization process) is small ([9, 20]). That is, we want to solve $(\mathcal{P}_d)$ for each λ_i belonging to a finite set of values and for a basis $(\psi_i)_{i\in I}$. This link between image compression and $(\mathcal{P}_d)$ might seem restrictive when we only consider a basis. It makes much more sense when we consider a redundant system of vectors $(\psi_i)_{i\in I}$. The use of redundant dictionaries has known a strong development in the past years, see [4, 17, 18, 3] for the most famous examples. In the context of dictionaries, we know that the length of the code for encoding $(\lambda_i)_{i\in I}$ is in general proportional to $\ell_0((\lambda_i)_{i\in I})$. The problem $(\mathcal{P}_d)$ therefore reads : minimize the codelength of the image while constraining a given level of accuracy of the coder. This is exactly the goal in image compression.

Finding an exact solution to $(\mathcal{P}_d)$ in large dimension (which is necessary in order to apply $(\mathcal{P}_d)$ to image compression) still remains a challenge. In fact, the methods described in [4, 18, 3] can be seen as heuristics approximating $(\mathcal{P}_d)$. The links between the performances of those heuristics and the performances of $(\mathcal{P}_d)$ is not completely clear. It is also a goal of the paper to provide a mean for comparing those algorithms.

1.2 Our contribution

In this paper, we estimate the ability of the model $(\mathcal{P}_d)$ to provide a sparse representation of data which follows a given distribution law. The distribution law is uniform in the θ-level set of a norm f_d :

$$\mathcal{L}_{f_d}(\theta) = \{w \in \mathbb{R}^N, f_d(w) \le \theta\}.$$

In order to do this we

- Give a precise (and non redundant) geometrical description of the sets

  $$\mathcal{I}^\tau(K) = \left\{d \in \mathbb{R}^N, \mathrm{val}(\mathcal{P}_d) \le K\right\},$$

  and

  $$\mathcal{D}^\tau(K) = \left\{d \in \mathbb{R}^N, \mathrm{val}(\mathcal{P}_d) = K\right\} \tag{5}$$

  where $\mathrm{val}(\mathcal{P}_d)$ denotes $\ell_0((\lambda_i)_{i\in I})$ for a solution $(\lambda_i)_{i\in I}$ of $(\mathcal{P}_d)$ and for $K = 0, \ldots, N$, $\tau > 0$. This is done in Theorem 1 and equation (69).

  Remark 1 *It is easy to see that* $\left\{\psi_i : \lambda_i \neq 0 \text{ for } (\lambda_i)_{i\in I} \text{ solving } (\mathcal{P}_d)\right\}$ *forms a set of linearly independent vectors. Therefore for all $d \in \mathbb{R}$ we will find a solution with at most N nonzero coefficients, even if the size of the dictionary is huge, $\#I \gg N$. So in this work we consider solutions with sparsity $K \le N$.*

- Once these sets are precisely described, we are able to bound (both from above and from below), their measure (more precisely the measure of their intersection with $\mathcal{L}_{f_d}(\theta)$). The difference between the upper and the lower bound is negligible when compared to $\left(\frac{\tau}{\theta}\right)^{N-K}$, when $\frac{\tau}{\theta}$ is mall enough. Moreover, these bounds show that the measures of $\mathcal{I}^\tau(K)\cap\mathcal{L}_{f_d}(\theta)$ and $\mathcal{D}^\tau(K)\cap\mathcal{L}_{f_d}(\theta)$ asymptotically behave like

$$\overline{\boldsymbol{C}}_K\theta^N\left(\frac{\tau}{\theta}\right)^{N-K},$$

as $\frac{\tau}{\theta}$ goes to 0.

The constants $\overline{\boldsymbol{C}}_K$ are defined in (44). They are made of the sum of constants C_V over all possible vector subspaces V of dimension K, spanned by elements of the dictionary $(\psi_i)_{i\in I}$. The constants C_V are built in Proposition 1 and Corollary 1. They have the form

$$C_V = \mathbb{L}^{N-K}\left(P_{V^\perp}\left(\mathcal{L}_{\|.\|}(1)\right)\right)\mathbb{L}^K\left(V\cap\mathcal{L}_{f_d}(1)\right),$$

where $P_{V^\perp}$ is the orthogonal projection onto the orthogonal complement of V, $\|.\|$ is the norm defining the data fidelity term in $(\mathcal{P}_d)$ and $\mathbb{L}^k(.)$ denotes the Lebesgue measure of a set living in $\mathbb{R}^k$.

- Once this is achieved, we easily obtain lower and upper bounds for $\mathbb{P}(\mathrm{val}(\mathcal{P}_d)\leq K)$, $\mathbb{P}(\mathrm{val}(\mathcal{P}_d)=K)$ when d is uniformly distributed in $\mathcal{L}_{f_d}(\theta)$ (see Section 6). They have the same characteristics as the bounds described above (modulo the disappearance of θ^N). In order to obtain sparse representations of the data, we should therefore tune the model (the norm $\|.\|$ and the dictionary $(\psi_i)_{i\in I}$) in order to obtain larger constants $\overline{\boldsymbol{C}}_K$.

  This result clearly shows that the model $(\mathcal{P}_d)$ benefits from several ingredient (which might not be present in other models promoting sparsity):

  - the sum defining $\overline{\boldsymbol{C}}_K$ is for all the possible vector subspaces of dimension K spanned by elements of the dictionary $(\psi_i)_{i\in I}$.
  - the term $\mathbb{L}^K\left(V\cap\mathcal{L}_{f_d}(1)\right)$ in in the constants C_V represents the measure of the whole set $V\cap\mathcal{L}_{f_d}(1)$.

- Finally we estimate $\mathbb{E}(\mathrm{val}(\mathcal{P}_d))$ and show that its asymptotic (when $\frac{\tau}{\theta}$ goes to 0) is governed by the constant $\overline{\boldsymbol{C}}_{N-1}$ (see Theorem 5). Increasing this constant therefore seems to be particularly important when building a model $(\mathcal{P}_d)$ (i.e. choosing $\|.\|$ and $(\psi_i)_{i\in I}$).

These results are illustrated in the context of particular choice for $\|.\|$ and for f_d in Section 7.

1.3 Relation to other evaluations of performance

Evaluating the performance of an optimization problem like $(\mathcal{P}_d)$ for the purpose of realizing nonlinear approximation is a very active firld of research. For a good survey of the problem we refer to [5].

In that field of research a variant of $(\mathcal{P}_d)$, named "best K-term approximation", is under study. It consists in looking for the best possible approximation of a datum $d\in\mathbb{R}^N$ using an expansion in $(\psi_i)_{i\in I}$ with K non-zero coordinates. The performance of the model is estimated using the quantity

$$\sigma_K(d) = \inf_{S\in\Sigma_K}\|d-S\|,$$

where Σ_K denotes the union of all the vector spaces of dimension K spanned by elements of $(\psi_i)_{i\in I}$, for $K = 0, \ldots, N$. Expressed with our notations, the typical object under consideration is[1]

$$\mathcal{A}^\alpha(C) = \bigcup_{K=1}^{N} \mathcal{D}^{\frac{C}{K^\alpha}}(K),$$

for $C > 0$ and $\alpha > 0$ and $\mathcal{D}^\tau(K)$ defined by (5). That is the data d obeying

$$\sigma_K(d) \le \frac{C}{K^\alpha} \quad \text{, for all } K = 1, \ldots, N.$$

The typical results obtained there take the form

$$\mathcal{A}^\alpha(C_1) \subset \mathcal{K}_\eta \subset \mathcal{A}^\alpha(C_2), \tag{6}$$

for $C_2 \ge C_1 > 0$ and the level set

$$\mathcal{K}_\eta = \{d \in \mathbb{R}^N, \|d\|_\eta \le 1\},$$

for a norm $\|.\|_\eta$ characterizing the regularity of d (again, the theory is in infinite dimensional vector spaces). This permits to estimate the number of coordinates which are needed to represent a datum d, if we know its regularity. Typically, the link between α and η says how good is the basis (or more generally a dictionary) at representing the data class.

The clear advantage of these results over ours is that they apply even if one only has a vague knowledge of the data distribution. For instance, any data distribution whose support is included in $\mathcal{K}_\eta$ does enjoy the decay $\frac{C_2}{K^\alpha}$. The inclusions in (6) need indeed to be true for the worse elements of $\mathcal{K}_\eta$ (even if they are rare). The counterpart of this advantage is that the constants C_1 and C_2 might be pessimistic.

Finally, as far as we know, the analysis proposed in Nonlinear approximation does not permit (today) to clearly assess the differences between $(\mathcal{P}_d)$ and its heuristics (in particular Basis Pursuit Denoising [3] and Orthogonal Matching Pursuit [18]). This is a clear advantage of the method for assessing model performances proposed in this paper. Indeed, similar analysis have already been conducted in [16, 13, 15] in the context of the compression scheme described in [14], Basis Pursuit Denoising and total variation regularization. (However, concerning the papers on Basis Pursuit Denoising and the total variation regularization, the results are stated for another asymptotic and the analysis partly needs to be rewritten in the proper context.)

1.4 Notations

For any function $f : \mathbb{R}^N \to \mathbb{R}$, and any $\theta \in \mathbb{R}$, the θ-level set of f is denoted by

$$\mathcal{L}_f(\theta) = \{w \in \mathbb{R}^N, f(w) \le \theta\}. \tag{7}$$

For any vector subspace V of $\mathbb{R}^N$, we denote P_V the orthogonal projection onto V and by $V^\perp$ the orthogonal complement of V in $\mathbb{R}^N$. To specify the dimension of V, we write $\dim(V)$. The Euclidean norm of an

[1] In Nonlinear approximation authors usually consider infinite dimensional spaces.

$u \in \mathbb{R}^N$ is systematically denoted by $\|u\|_2$. The notation $\|u\|$ is devoted to a general norm on $\mathbb{R}^N$. For any integer $K > 0$, the Lebesgue measure on $\mathbb{R}^K$ is systematically denoted by $\mathbb{L}^K(.)$, whereas I_K stands for the $K \times K$ identity matrix. We write $\mathbb{P}(.)$ for probability and $\mathbb{E}(.)$ for expectation.

As usually, we write $o(t)$ for a function satisfying $\lim_{t\to 0} \frac{o(t)}{t} = 0$.

For any $d \in \mathbb{R}^N$, we denote $\mathrm{val}(\mathcal{P}_d)$ the value of the minimum in $(\mathcal{P}_d)$—i.e. $\ell_0\left((\lambda_i)_{i\in I}\right)$ for $(\lambda_i)_{i\in I}$ solving $(\mathcal{P}_d)$.

2 Measuring bounded cylinder-like subsets of $\mathbb{R}^N$

2.1 Preliminary results

Below we give several statements that will be used many times in the rest of the work.

Lemma 1 *For any vector subspace $V \subset \mathbb{R}^N$ and any norm $\|.\|$ on $\mathbb{R}^N$, define the application*

$$\begin{aligned} h : V^\perp &\to \mathbb{R} \\ u &\to h(u) \stackrel{\text{def}}{=} \inf\left\{t \geq 0 : \frac{u}{t} \in P_{V^\perp}\left(\mathcal{L}_{\|.\|}(1)\right)\right\}. \end{aligned} \tag{8}$$

Then the following holds:

(i) For any $\tau \geq 0$, we have

$$\mathcal{L}_h(\tau) = P_{V^\perp}\left(\mathcal{L}_{\|.\|}(\tau)\right). \tag{9}$$

(ii) The application h in (8) *is a norm on $V^\perp$.*

(iii) For any norm f_d on $\mathbb{R}^N$, let $\delta_1 > 0$, $\delta_2 > 0$ and $\overline{\Delta}$ be some constants satisfying

$$w \in \mathbb{R}^N \quad \Rightarrow \quad f_d(w) \leq \delta_1 \|w\|_2 \quad \text{and} \quad \|w\|_2 \leq \delta_2 \|w\|, \tag{10}$$

$$\overline{\Delta} \stackrel{\text{def}}{=} \delta_1 \delta_2 \tag{11}$$

The constants δ_1, δ_2 and $\overline{\Delta} > 0$ are independent of V and we have

$$f_d(u) \leq \overline{\Delta} h(u), \qquad \forall u \in V^\perp, \tag{12}$$

$$\|u\|_2 \leq \delta_2 h(u), \qquad \forall u \in V^\perp. \tag{13}$$

Remark 2 *The constants in (10) come from the fact that all norms on a finite-dimensional space are equivalent. In practice we will choose the smallest constants satisfying these inequalities.*

Proof. The case $V = \{0\}$ is trivial (we obtain $h = \|.\|$) and we further assume that $\dim(V) \geq 1$.

Assertion (i). The set $P_{V^\perp}(\mathcal{L}_{\|.\|}(1))$ is convex since $\|.\|$ is a norm and $P_{V^\perp}$ is linear. Moreover, the origin 0 belongs to its interior. Indeed, there is $\varepsilon > 0$ such that if $w \in \mathbb{R}^N$ satisfies $\|w\|_2 < \varepsilon$, then $\|w\| < 1$. Consequently $0 \in \mathrm{Int}\big(\mathcal{L}_{\|.\|_2}(\varepsilon)\big) \subset \mathcal{L}_{\|.\|}(1)$. Using that $\|.\|_2$ is rotationally invariant and that $P_{V^\perp}$ is a contraction, we deduce that $0 \in \mathrm{Int}\big(P_{V^\perp}(\mathcal{L}_{\|.\|_2}(\varepsilon))\big) \subset P_{V^\perp}(\mathcal{L}_{\|.\|}(1))$. Then the application $h : V^\perp \to \mathbb{R}$ in (8) is the usual Minkowski functional of $P_{V^\perp}\left(\mathcal{L}_{\|.\|}(1)\right)$, as defined and commented in [12, p.131]. Since $P_{V^\perp}\left(\mathcal{L}_{\|.\|}(1)\right)$ is closed, we have

$$P_{V^\perp}\left(\mathcal{L}_{\|.\|}(1)\right) = \left\{u \in V^\perp : h(u) \leq 1\right\}.$$

Using that the Minkowski functional is positively homogeneous—i.e.

$$h(\tau u) = \tau h(u), \quad \forall \tau > 0,$$

lead to (9).

Assertion (ii). For h to be a norm, we have to show that the latter property holds for any $\lambda \in \mathbb{R}$ (i.e. that h is symmetric with respect to the origin). It is true since, for any $\lambda \in \mathbb{R}$

$$\begin{aligned} h(\lambda u) &= \inf\left\{t \geq 0 : \lambda u \in P_{V^\perp}\left(\mathcal{L}_{\|.\|}(t)\right)\right\} \\ &= \inf\left\{t \geq 0 : u \in P_{V^\perp}\left(\mathcal{L}_{\|.\|}(\frac{t}{|\lambda|})\right)\right\} \\ &= |\lambda| \inf\left\{t \geq 0 : u \in P_{V^\perp}\left(\mathcal{L}_{\|.\|}(t)\right)\right\} \qquad \text{(writing } t \text{ for } t/|\lambda|) \\ &= |\lambda|\ h(u), \end{aligned}$$

where we use the facts that $P_{V^\perp}$ is linear and that $\|.\|$ is a norm. It is well known that the Minkowski functional is non negative, finite, and satisfies[2] $h(u+v) \leq h(u) + h(v)$ for any $u, v \in V^\perp$.

Finally, since $\mathcal{L}_h(0) = P_{V^\perp}\left(\mathcal{L}_{\|.\|}(0)\right) = \{0\}$,

$$h(u) = 0 \quad \Leftrightarrow \quad u = 0.$$

Consequently, h defines a norm on $V^\perp$.

Assertion (iii). Let us first remark that

$$\mathcal{L}_{\|.\|}(1) \subset \mathcal{L}_{\|.\|_2}(\delta_2) \subset \mathcal{L}_{f_d}(\delta_1\delta_2) = \mathcal{L}_{f_d}(\overline{\Delta}),$$

where δ_1 and δ_2 are defined in the proposition. Using that $\|.\|_2$ is rotationally invariant, we have

$$\begin{aligned} \mathcal{L}_h(1) = P_{V^\perp}\left(\mathcal{L}_{\|.\|}(1)\right) &\subset P_{V^\perp}\left(\mathcal{L}_{\|.\|_2}(\delta_2)\right) = \mathcal{L}_{\|.\|_2}(\delta_2) \cap V^\perp \\ &\subset \mathcal{L}_{f_d}(\delta_1\delta_2) \cap V^\perp = \mathcal{L}_{f_d}(\overline{\Delta}) \cap V^\perp. \end{aligned}$$

We will prove (12) and (13) jointly. To this end let us consider a norm g on $\mathbb{R}^N$ and $\delta > 0$ such that

$$P_{V^\perp}\left(\mathcal{L}_{\|.\|}(1)\right) \subset \mathcal{L}_g(\delta) \cap V^\perp. \tag{14}$$

[2] For completeness, we give the details:

$$\begin{aligned} h(u+v) &= \inf\left\{t \geq 0 : (u+v) \in P_{V^\perp}\left(\mathcal{L}_{\|.\|}(t)\right)\right\} \\ &\leq \inf\left\{t \geq 0 : u \in P_{V^\perp}\left(\mathcal{L}_{\|.\|}(t)\right)\right\} + \inf\left\{t \geq 0 : v \in P_{V^\perp}\left(\mathcal{L}_{\|.\|}(t)\right)\right\} = h(u) + h(v). \end{aligned}$$

Using that each norm can be expressed as a Minkowski functional, for any $u \in V^\perp$ we can write down the following:

$$\begin{aligned}
g(u) &= \inf\{t \geq 0 : g(\frac{u}{t}) \leq 1\} \\
&= \inf\{t \geq 0 : g(\frac{\delta}{t}u) \leq \delta\} \\
&= \delta \inf\{t \geq 0 : g(\frac{u}{t}) \leq \delta\} \qquad\qquad (\text{write } t \text{ for } \tfrac{t}{\delta}) \\
&= \delta \inf\{t \geq 0 : \frac{u}{t} \in \mathcal{L}_g(\delta)\} \\
&\leq \delta \inf\{t \geq 0 : \frac{u}{t} \in P_{V^\perp}\left(\mathcal{L}_{\|.\|}(1)\}\right) \qquad (15)\\
&\leq \delta\ h(u),
\end{aligned}$$

where the inequality in (15) comes from (14).

If we identify g with f_d and δ with $\overline{\Delta}$, we obtain (12). Similarly, identifying g with $\|.\|_2$ and δ with δ_2 yields (13). This concludes the proof. □

The next proposition addresses sets of $\mathbb{R}^N$ bounded with the aid of f_d.

Proposition 1 *For any vector subspace V of $\mathbb{R}^N$, any norm $\|.\|$ on $\mathbb{R}^N$ and any $\tau > 0$, define*

$$V^\tau = V + P_{V^\perp}\left(\mathcal{L}_{\|.\|}(\tau)\right). \qquad (16)$$

Then the following hold:

(i) V^τ is closed and measurable;

(ii) Let f_d be any norm on $\mathbb{R}^N$, $h : V^\perp \to \mathbb{R}$ the norm defined in Lemma 1, $K = \dim(V)$ and δ_V be any constant such that

$$f_d(u) \leq \delta_V h(u), \quad \forall u \in V^\perp. \qquad (17)$$

If $\theta \geq \delta_V \tau$, then

$$C\tau^{N-K}(\theta - \delta_V \tau)^K \leq \mathbb{L}^N\left(V^\tau \cap \mathcal{L}_{f_d}(\theta)\right) \leq C\tau^{N-K}(\theta + \delta_V \tau)^K, \qquad (18)$$

where

$$C = \mathbb{L}^{N-K}\left(P_{V^\perp}\left(\mathcal{L}_{\|.\|}(1)\right)\right)\mathbb{L}^K\left(V \cap \mathcal{L}_{f_d}(1)\right) \in (0, +\infty). \qquad (19)$$

Remark 3 *Using Lemma 1, the condition in (17) holds for any $\delta_V \geq \delta_V^*$ with $\delta_V^* \in [0, \overline{\Delta}]$, where $\overline{\Delta}$ is given in* (11). *Let us emphasize that δ_V may depend on V (which explains the letter "V" in index). The proposition clearly holds if we take $\delta_V = \overline{\Delta}$—the constant of Lemma 1, assertion (iii), which is independent of the choice of V.*

Observe that C is a positive, finite constant that depends only on V, $\|.\|$ and f_d.

Remark 4 *An important consequence of this proposition is that asymptotically*

$$\mathbb{L}^N\left(V^\tau \cap \mathcal{L}_{f_d}(\theta)\right) = C\theta^N \left(\frac{\tau}{\theta}\right)^{N-K} + \theta^N o\left(\left(\frac{\tau}{\theta}\right)^{N-K}\right) \quad \textit{if} \quad \frac{\tau}{\theta} \to 0.$$

Proof. The sets V and $P_{V^\perp}\left(\mathcal{L}_{\|.\|}(\tau)\right)$ are closed. Moreover, V and $P_{V^\perp}\left(\mathcal{L}_{\|.\|}(\tau)\right)$ are orthogonal. Therefore V^τ is closed. As a consequence V^τ is a Borel set and is Lebesgue measurable.

Since the restriction of f_d to $V^\perp$ is a norm on $V^\perp$, there exists δ_V such that (see Remark 2)

$$f_d(u) \leq \delta_V h(u), \quad \forall u \in V^\perp, \tag{20}$$

where h is given in (9) in Lemma 1. By (12) in Lemma 1, such a δ_V exists in $[0, \overline{\Delta}]$. To simplify the notations, in the rest of the proof we will write δ for δ_V.

For any $u \in V^\perp$ and $v \in V$, using (20) we have

$$f_d(v) - \delta h(u) \leq f_d(v) - f_d(u) \leq f_d(u+v) \leq f_d(v) + f_d(u) \leq f_d(v) + \delta h(u)$$

In particular, for $h(u) \leq \tau$, we get

$$f_d(v) - \delta\tau \leq f_d(u+v) \leq f_d(v) + \delta\tau. \tag{21}$$

As required in assertion (ii), we have $\theta - \delta\tau \geq 0$. If in addition $v \in V$ is such that $f_d(v) \leq \theta - \delta\tau$, then $f_d(u+v) \leq \theta$. Noticing that

$$\mathcal{L}_{f_d}(\theta) = \left\{u + v : (u,v) \in (V^\perp \times V),\ f_d(u+v) \leq \theta\right\},$$

this implies that

$$B_0 \stackrel{\text{def}}{=} \left\{u + v : (u,v) \in (V^\perp \times V),\ h(u) \leq \tau,\ f_d(v) \leq \theta - \delta\tau\right\} \ \subseteq\ V^\tau \cap \mathcal{L}_{f_d}(\theta).$$

Using that $f_d(u+v) \leq \theta$ (see the set we wish to measure in (18)), then the left-hand side of (21) shows that $f_d(v) \leq \theta + \delta\tau$, hence

$$B_1 \stackrel{\text{def}}{=} \left\{u + v : (u,v) \in (V^\perp \times V),\ h(u) \leq \tau,\ f_d(v) \leq \theta + \delta\tau\right\} \ \supseteq\ V^\tau \cap \mathcal{L}_{f_d}(\theta).$$

Consider the pair of applications

$$\begin{array}{rcl} \varphi_0 : \mathcal{L}_h(1) \times (V \cap \mathcal{L}_{f_d}(1)) & \to & \mathbb{R}^N \\ (u,v) & \to & \tau u + (\theta - \delta\tau) v \end{array}$$

and

$$\begin{array}{rcl} \varphi_1 : \mathcal{L}_h(1) \times (V \cap \mathcal{L}_{f_d}(1)) & \to & \mathbb{R}^N \\ (u,v) & \to & \tau u + (\theta + \delta\tau) v \end{array}$$

Clearly, φ_i is a Lipschitz homeomorphism satisfying $\varphi_i\Big(\mathcal{L}_h(1) \times \left(V \cap \mathcal{L}_{f_d}(1)\right)\Big) = B_i$ for $i \in \{0,\ 1\}$. Moreover, we have

$$D\varphi_0 = \begin{bmatrix} \tau I_{N-K} & 0 \\ 0 & (\theta - \delta\tau) I_K \end{bmatrix} \quad \text{and} \quad D\varphi_1 = \begin{bmatrix} \tau I_{N-K} & 0 \\ 0 & (\theta + \delta\tau) I_K \end{bmatrix}.$$

Then $\mathbb{L}^N\left(B_i\right)$ can be computed using (see [7] for details)

$$\mathbb{L}^N\left(B_i\right) = \int_{u\in\mathcal{L}_h(1)}\int_{v\in V\cap\mathcal{L}_{f_d}(1)} [\![\varphi_i]\!] dv du,$$

where $[\![\varphi_i]\!]$ is the Jacobian of φ_i, for $i=0$ or $i=1$. In particular,

$$\begin{aligned}
[\![\varphi_0]\!] &= \det\left(D\varphi_0\right) = \tau^{N-K}(\theta-\delta\tau)^K, \\
[\![\varphi_1]\!] &= \det\left(D\varphi_1\right) = \tau^{N-K}(\theta+\delta\tau)^K.
\end{aligned}$$

It follows that

$$\mathbb{L}^N\left(B_0\right) = C\tau^{N-K}(\theta-\delta\tau)^K \quad \text{and} \quad \mathbb{L}^N\left(B_1\right) = C\tau^{N-K}(\theta+\delta\tau)^K$$

where the constant

$$\begin{aligned}
C &= \int_{\mathcal{L}_h(1)} du \int_{V\cap\mathcal{L}_{f_d}(1)} dv \\
&= \mathbb{L}^{N-K}\left(P_{V^\perp}\left(\mathcal{L}_{\|.\|}(1)\right)\right)\mathbb{L}^K\left(V\cap\mathcal{L}_{f_d}(1)\right).
\end{aligned}$$

Clearly C is positive and finite. Using the inclusion $B_0 \subseteq V^\tau \cap \mathcal{L}_{f_d}(\theta) \subseteq B_1$ shows that

$$C\tau^{N-K}(\theta-\delta\tau)^K \leq \mathbb{L}^N\left(V^\tau\cap\mathcal{L}_{f_d}(\theta)\right) \leq C\tau^{N-K}(\theta+\delta\tau)^K.$$

The proof is complete. $\square$

2.2 Sets built from a dictionary

With every $J\subset I$, we associate the vector subspace $\mathcal{T}_J$ defined below:

$$\mathcal{T}_J \stackrel{\text{def}}{=} \operatorname{span}\left((\psi_j)_{j\in J}\right), \tag{22}$$

along with the convention $\operatorname{span}(\emptyset)=\{0\}$. Given an arbitrary $\tau>0$, we introduce the subset of $\mathbb{R}^N$

$$\mathcal{T}_J^\tau \stackrel{\text{def}}{=} \mathcal{T}_J + P_{\mathcal{T}_J^\perp}\left(\mathcal{L}_{\|.\|}(\tau)\right), \tag{23}$$

where we recall that $\mathcal{T}_J^\perp$ is the orthogonal complement of $\mathcal{T}_J$ in $\mathbb{R}^N$ and $\|.\|$ is any norm on $\mathbb{R}^N$. These notations are constantly used in what follows.

The next assertion is a direct consequence of Proposition 1. The proposition is illustrated on Figure 1.

Corollary 1 *For any $J\subset I$ (including $J=\emptyset$), any norm $\|.\|$ and any $\tau>0$ the following hold:*

(i) $\mathcal{T}_J^\tau$ is closed and measurable;

(ii) Let f_d be any norm on $\mathbb{R}^N$ and $K \stackrel{\text{def}}{=} \dim(\mathcal{T}_J)$. Then there exists $\delta_J\in[0,\overline{\Delta}]$ (where $\overline{\Delta}$ is given in Lemma 1(iii)) such that for $\theta\geq\delta_J\tau$ we have

$$C_J\tau^{N-K}(\theta-\delta_J\tau)^K \;\leq\; \mathbb{L}^N\left(\mathcal{T}_J^\tau\cap\mathcal{L}_{f_d}(\theta)\right) \;\leq\; C_J\tau^{N-K}(\theta+\delta_J\tau)^K, \tag{24}$$

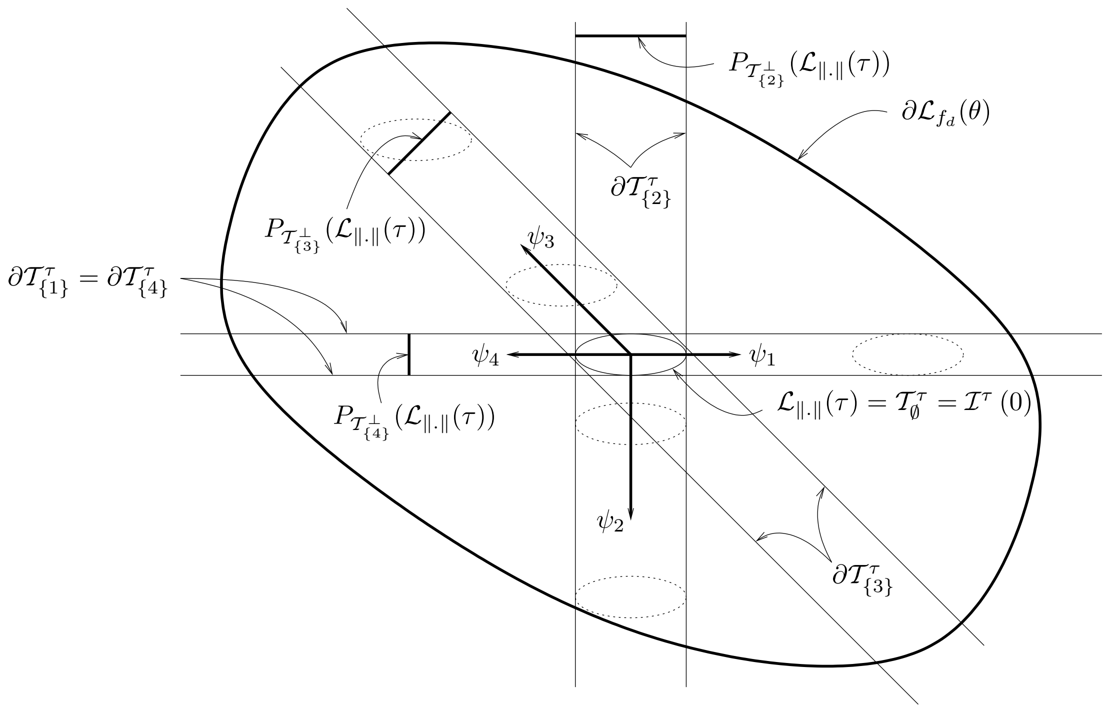

Figure 1: Example in dimension 2. Let the dictionary read $\{\psi_1, \psi_2, \psi_3, \psi_4\}$. On the drawing, the sets $P_{\mathcal{T}_{\{i\}}^{\perp}}(\mathcal{L}_{\|.\|}(\tau))$, for $i = 2, 3, 4$, are shifted by an element of $\mathcal{T}_{\{i\}}$. The dotted sets represent translations of $\mathcal{L}_{\|.\|}(\tau)$. The set-valued function $\mathcal{I}^{\tau}()$, as presented in (39) and Proposition 3, gives rise to the following situations: $\mathcal{I}^{\tau}(0) = \mathcal{L}_{\|.\|}(\tau) = \mathcal{T}_{\emptyset}^{\tau}$, $\mathcal{I}^{\tau}(1) = \mathcal{T}_{\{1\}}^{\tau} \cup \mathcal{T}_{\{2\}}^{\tau} \cup \mathcal{T}_{\{3\}}^{\tau}$ and $\mathcal{I}^{\tau}(2) = \mathbb{R}^2 = \mathcal{T}_{\{1,2\}}^{\tau} = \mathcal{T}_{\{2,3\}}^{\tau} = \ldots$ The symbol ∂ is used to denote the boundaries of the sets.

where

$$C_J = \mathbb{L}^{N-K}\left(P_{\mathcal{T}_J^\perp}\left(\mathcal{L}_{\|.\|}(1)\right)\right)\,\mathbb{L}^{K}\left(\mathcal{T}_J \cap \mathcal{L}_{f_d}(1)\right) \in (0,+\infty). \tag{25}$$

Proof. The corollary is a direct consequence of Proposition 1. Notice that we now write δ_J for the constant $\delta_{\mathcal{T}_J}$ in Lemma 1.

□

It can be useful to remind that $\overline{\Delta}$ is defined in Lemma 1 and only depends on $\|.\|$ and f_d.

A more friendly expression for $\mathcal{T}_J^\tau$ is provided by the lemma below. Again, the lemma is illustrated on Figure 1.

Lemma 2 *For any $J \subset I$ (including $J = \emptyset$), any norm $\|.\|$ and $\tau > 0$ let $\mathcal{T}_J^\tau$ be defined by (23). Then*

$$\mathcal{T}_J^\tau = \mathcal{T}_J + \mathcal{L}_{\|.\|}(\tau).$$

Proof. The case $J = \emptyset$ is trivial because of the convention $\mathrm{span}(\emptyset) = \{0\}$. Consider next that J is nonempty. Let $w \in \mathcal{T}_J^\tau$, then w admits a unique decomposition as

$$w = v + u \quad \text{where} \quad v \in \mathcal{T}_J \quad \text{and} \quad u \in \mathcal{T}_J^\perp.$$

If $\|u\| \le \tau$ then clearly $w \in \mathcal{T}_J + \mathcal{L}_{\|.\|}(\tau)$. Consider next that $\|u\| > \tau$. From the definition of $\mathcal{T}_J^\tau$, there exists $w_u \in \mathcal{L}_{\|.\|}(\tau)$ such that $P_{\mathcal{T}_J^\perp}(w_u) = u$. Noticing that $u - w_u = P_{\mathcal{T}_J^\perp}(w_u) - w_u \in \mathcal{T}_J$ and that $v + u - w_u \in \mathcal{T}_J$, we can see that

$$\begin{aligned} w &= (v + u - w_u) + w_u \\ &\in \quad \mathcal{T}_J \quad + \mathcal{L}_{\|.\|}(\tau). \end{aligned}$$

Conversely, let $w \in \mathcal{T}_J + \mathcal{L}_{\|.\|}(\tau)$. Then

$$w = v_1 + v \quad \text{where} \quad v_1 \in \mathcal{T}_J \quad \text{and} \quad v \in \mathcal{L}_{\|.\|}(\tau).$$

Furthermore, v has a unique decomposition of the form

$$v = v_2 + u \quad \text{where} \quad v_2 \in \mathcal{T}_J \quad \text{and} \quad u \in \mathcal{T}_J^\perp.$$

In particular,

$$u = P_{\mathcal{T}_J^\perp}(v) \in P_{\mathcal{T}_J^\perp}\left(\mathcal{L}_{\|.\|}(\tau)\right)$$

Combining this with the fact that $v_1 + v_2 \in \mathcal{T}_J$ shows that $w = (v_1 + v_2) + u \in \mathcal{T}_J^\tau$. □

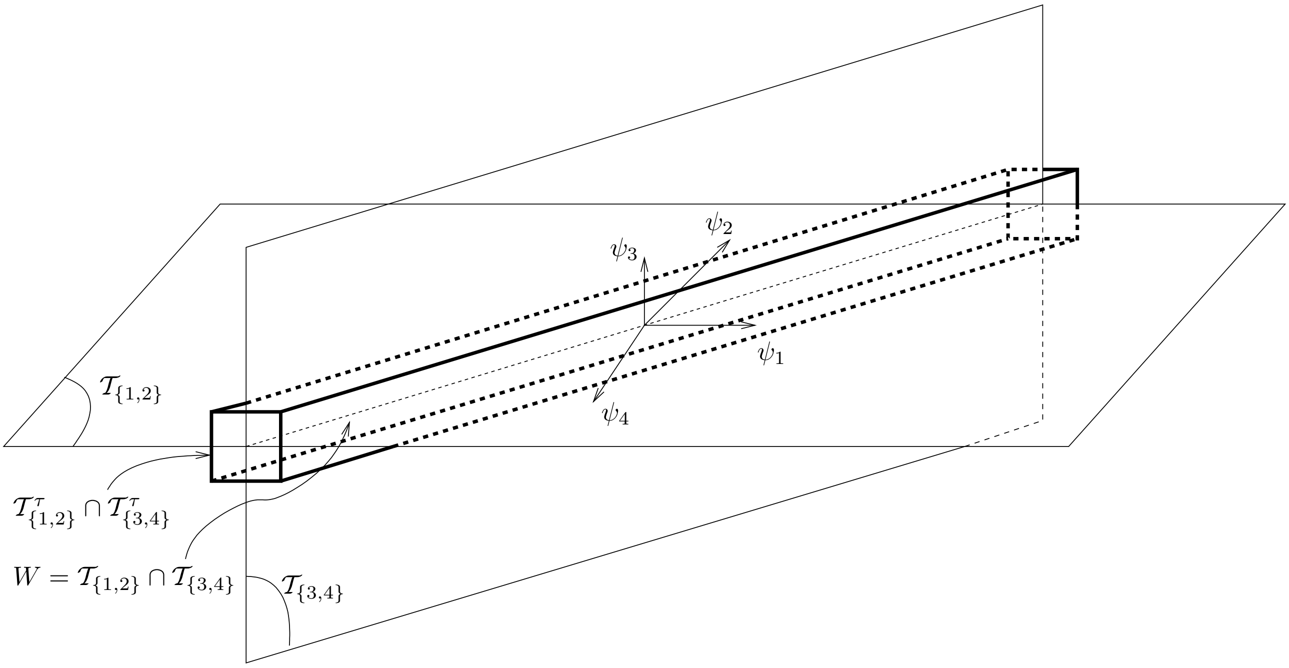

Figure 2: Example of an intersection in dimension 3. $\mathcal{T}^\tau_{\{1,2\}}$ is in between to planes, parallel to $\mathcal{T}_{\{1,2\}}$. Same remark for $\mathcal{T}^\tau_{\{3,4\}}$. The set $\mathcal{T}^\tau_{\{1,2\}} \cap \mathcal{T}^\tau_{\{3,4\}}$ is of the form $W + P_{W^\perp}\mathcal{L}_{\tilde{g}}(\tau)$, where $\tilde{g}$ is a norm and for $W = \mathcal{T}_{\{1,2\}} \cap \mathcal{T}_{\{3,4\}}$. We also have $\dim(\mathcal{T}_{\{1,2\}} \cap \mathcal{T}_{\{3,4\}}) < dim(\mathcal{T}_{\{1,2\}}) = dim(\mathcal{T}_{\{3,4\}})$.

3 The intersection of two cylinder-like subsets is small

This section is devoted to prove quite an intuitive result on the estimate of the intersection of two sets $\mathcal{T}^\tau_J$. It uses all notations introduced in §2.2 and is illustrated on Figure 2.

Proposition 2 *Let $J_1 \subset I$ and $J_2 \subset I$ be such that $\mathcal{T}_{J_1} \neq \mathcal{T}_{J_2}$ and $\dim(\mathcal{T}_{J_1}) = \dim(\mathcal{T}_{J_2}) \stackrel{\text{def}}{=} K$. Let $\tau > 0$ and $\theta > 0$. Then the set given below*

$$\mathcal{T}^\tau_{J_1} \cap \mathcal{T}^\tau_{J_2} \cap \mathcal{L}_{f_d}(\theta) \tag{26}$$

is closed and measurable. Moreover, there is a constant $\delta_{J_1,J_2} \in [0, 3\overline{\Delta}]$ (where $\overline{\Delta}$ is given in Lemma 1(iii)) such that for $\theta \geq \delta_{J_1,J_2}\tau$ we have

$$\mathbb{L}^N\left(\mathcal{T}^\tau_{J_1} \cap \mathcal{T}^\tau_{J_2} \cap \mathcal{L}_{f_d}(\theta)\right) \leq Q_{J_1,J_2}\tau^{N-k}(\theta + \delta_{J_1,J_2}\tau)^k, \qquad k = \dim\left(\mathcal{T}_{J_1} \cap \mathcal{T}_{J_2}\right),$$

where the constant Q_{J_1,J_2} reads

$$Q_{J_1,J_2} \stackrel{\text{def}}{=} \mathbb{L}^{N-k}\left(W^\perp \cap \mathcal{L}_{\|.\|_2}(2\delta_2)\right)\mathbb{L}^k\left(W \cap \mathcal{L}_{f_d}(1)\right) \quad \textit{for} \quad W \stackrel{\text{def}}{=} \mathcal{T}_{J_1} \cap \mathcal{T}_{J_2}. \tag{27}$$

Notice that Q_{J_1,J_2} depends only on $(\psi_j)_{j\in J_1}$ and $(\psi_j)_{j\in J_2}$, and the norms $\|.\|$ and f_d. A tighter bound can be found in the proof of the proposition (see equation (37)). The bound is expressed in terms of a norm $\tilde{g}$ constructed there.

Remark 5 *Since $k = \dim W \leq K-1$, we have the following asymptotical result:*

$$\begin{aligned}
\mathbb{L}^N\left(\mathcal{T}^\tau_{J_1} \cap \mathcal{T}^\tau_{J_2} \cap \mathcal{L}_{f_d}(\theta)\right) &\leq Q_{J_1,J_2}\ \theta^N \left(\frac{\tau}{\theta}\right)^{N-k} \left(1+\delta_{J_1,J_2}\frac{\tau}{\theta}\right)^k \\
&= Q_{J_1,J_2}\ \theta^N \left(\frac{\tau}{\theta}\right)^{N-k} + o\left(\left(\frac{\tau}{\theta}\right)^{N-k}\right) \qquad \text{as } \frac{\tau}{\theta} \to 0 \\
&= \theta^N o\left(\left(\frac{\tau}{\theta}\right)^{N-K}\right) \qquad \text{as } \frac{\tau}{\theta} \to 0.
\end{aligned}$$

Proof. The subset in (26) is closed and measurable, as being a finite intersection of closed measurable sets.

Let

$$h_1 : \mathcal{T}^\perp_{J_1} \to \mathbb{R} \quad \text{and} \quad h_2 : \mathcal{T}^\perp_{J_2} \to \mathbb{R}$$

be the norms exhibited in Lemma 1—see equation (8)—such that for any $\tau \geq 0$,

$$\mathcal{L}_{h_1}(\tau) = P_{\mathcal{T}_{J_1^\perp}}\left(\mathcal{L}_{\|\cdot\|}(\tau)\right) \quad \text{and} \quad \mathcal{L}_{h_2}(\tau) = P_{\mathcal{T}_{J_2^\perp}}\left(\mathcal{L}_{\|\cdot\|}(\tau)\right).$$

Reminding that by definition

$$W = \mathcal{T}_{J_1} \cap \mathcal{T}_{J_2},$$

De Morgan's law shows that

$$W^\perp = \mathcal{T}^\perp_{J_1} + \mathcal{T}^\perp_{J_2}.$$

Below we express the latter sum as a direct sum of subspaces:

$$\begin{aligned}
W^\perp = \left(\mathcal{T}^\perp_{J_1} \cap \mathcal{T}^\perp_{J_2}\right) \quad &\oplus \quad \left(\mathcal{T}^\perp_{J_1} \cap \mathcal{T}_{J_2}\right) \\
&\oplus \quad \left(\mathcal{T}_{J_1} \cap \mathcal{T}^\perp_{J_2}\right).
\end{aligned} \tag{28}$$

Notice that we have

$$\begin{aligned}
\mathcal{T}^\perp_{J_1} &= \left(\mathcal{T}^\perp_{J_1} \cap \mathcal{T}^\perp_{J_2}\right) \oplus \left(\mathcal{T}^\perp_{J_1} \cap \mathcal{T}_{J_2}\right), \\
\mathcal{T}^\perp_{J_2} &= \left(\mathcal{T}^\perp_{J_1} \cap \mathcal{T}^\perp_{J_2}\right) \oplus \left(\mathcal{T}_{J_1} \cap \mathcal{T}^\perp_{J_2}\right),
\end{aligned} \tag{29}$$

as well as

$$\begin{aligned}
\mathcal{T}_{J_1} &= W \oplus \left(\mathcal{T}_{J_1} \cap \mathcal{T}^\perp_{J_2}\right), \\
\mathcal{T}_{J_2} &= W \oplus \left(\mathcal{T}^\perp_{J_1} \cap \mathcal{T}_{J_2}\right).
\end{aligned} \tag{30}$$

¿From (28), any $u \in W^\perp$ has a unique decomposition as

$$u = u_1 + u_2 + u_3 \qquad \text{where} \qquad \begin{array}{l} u_1 \in \mathcal{T}^\perp_{J_1} \cap \mathcal{T}^\perp_{J_2} \\ u_2 \in \mathcal{T}^\perp_{J_1} \cap \mathcal{T}_{J_2} \\ u_3 \in \mathcal{T}_{J_1} \cap \mathcal{T}^\perp_{J_2} \end{array} \tag{31}$$

Using these notations, we introduce the following function:

$$\begin{aligned}
g : W^\perp &\to \mathbb{R} \\
u &\to g(u) = \sup\left\{h_1(u_1+u_2), h_2(u_1+u_3)\right\}.
\end{aligned} \tag{32}$$

In the next lines we show that g is a norm on $W^\perp$:

- h_1 and h_2 being norms, $g(\lambda u) = |\lambda| g(u)$, for all $\lambda \in \mathbb{R}$;
- if $g(u) = 0$ then $u_1 + u_2 = u_1 + u_3 = 0$; noticing that $u_1 \perp u_2$ and that $u_1 \perp u_3$ yields $u = 0$;
- for $u \in W^\perp$ and $v \in W^\perp$ (both decomposed according to (31)),

$$\begin{aligned} g(u+v) &= \sup\big\{h_1(u_1+u_2+v_1+v_2), h_2(u_1+u_3+v_1+v_3)\big\} \\ &\le \sup\big\{h_1(u_1+u_2)+h_1(v_1+v_2), h_2(u_1+u_3)+h_2(v_1+v_3)\big\} \\ &\le \sup\big\{h_1(u_1+u_2), h_2(u_1+u_3)\big\} + \sup\big\{h_1(v_1+v_2), h_2(v_1+v_3)\big\} \\ &= g(u)+g(v). \end{aligned}$$

Furthermore, g can be extended to a norm $\tilde{g}$ on $\mathbb{R}^N$ such that $\forall u \in W^\perp$, we have $\tilde{g}(u) = g(u)$ and

$$\mathcal{L}_g(\tau) = P_{W^\perp}\left(\mathcal{L}_{\tilde{g}}(\tau)\right), \qquad \forall \tau > 0. \tag{33}$$

Let us then define

$$\begin{aligned} W^\tau &= W + P_{W^\perp}\left(\mathcal{L}_{\tilde{g}}(\tau)\right) \\ &= \Big\{w + u : (u,w) \in (W^\perp \times W),\ g(u) \le \tau\Big\}. \end{aligned} \tag{34}$$

We are going to show that $\left(\mathcal{T}_{J_1}^\tau \cap \mathcal{T}_{J_2}^\tau\right) \subset W^\tau$. In order to do so, we consider an arbitrary

$$v \in \mathcal{T}_{J_1}^\tau \cap \mathcal{T}_{J_2}^\tau. \tag{35}$$

It admits a unique decomposition of the form

$$v = w + u_1 + u_2 + u_3,$$

where $w \in W$, and $u_1,\ u_2$ and u_3 are decomposed according to (31). The latter, combined with (29) and (30) shows that

$$\begin{aligned} u_1 + u_2 \in \mathcal{T}_{J_1}^\perp \quad &\text{and} \quad w + u_3 \in \mathcal{T}_{J_1}, \\ u_1 + u_3 \in \mathcal{T}_{J_2}^\perp \quad &\text{and} \quad w + u_2 \in \mathcal{T}_{J_2}. \end{aligned}$$

The inclusions given above, combined with (35), show that

$$h_1(u_1+u_2) \le \tau \quad \text{and} \quad h_2(u_1+u_3) \le \tau.$$

By the definition of g in (31)-(32), the inequalities given above imply that $g(u) \le \tau$. Combining this with the definition of W^τ in (34) entails that $v \in W^\tau$. Consequently,

$$\left(\mathcal{T}_{J_1}^\tau \cap \mathcal{T}_{J_2}^\tau\right) \subset W^\tau \quad \text{and} \quad \left(\mathcal{T}_{J_1}^\tau \cap \mathcal{T}_{J_2}^\tau \cap \mathcal{L}_{f_d}(\theta)\right) \ \subset\ \left(W^\tau \cap \mathcal{L}_{f_d}(\theta)\right).$$

It follows that

$$\mathbb{L}^N\left(\mathcal{T}_{J_1}^\tau \cap \mathcal{T}_{J_2}^\tau \cap \mathcal{L}_{f_d}(\theta)\right) \le \mathbb{L}^N\left(W^\tau \cap \mathcal{L}_{f_d}(\theta)\right).$$

Applying now the right-hand side of (18) in Proposition 1 with W^τ in place of V^τ and taking δ_{J_1,J_2} such that

$$f_d(u) \le \delta_{J_1,J_2} g(u), \quad \forall u \in W^\perp, \tag{36}$$

leads to

$$\mathbb{L}^N\left(W^\tau \cap \mathcal{L}_{f_d}(\theta)\right) \le Q'_{J_1,J_2} \tau^{N-k}(\theta + \delta_{J_1,J_2}\tau)^k,$$

where it is easy to see that

$$Q'_{J_1,J_2} = \mathbb{L}^{N-k}\left(P_{W^\perp}\left(\mathcal{L}_{\tilde{g}}(\tau)\right)\right)\mathbb{L}^k\left(W \cap \mathcal{L}_{f_d}(1)\right) = \mathbb{L}^{N-k}\left(\mathcal{L}_g(1)\right)\mathbb{L}^k\left(W \cap \mathcal{L}_{f_d}(1)\right). \tag{37}$$

In order to obtain (27), we are going to show that $\mathcal{L}_g(1) \subset \left(\mathcal{L}_{\|.\|_2}(2\delta_2) \cap W^\perp\right)$. Using Lemma 1 (ii), if $u \in W^\perp$ is decomposed according to (31), we obtain

$$\begin{aligned}
\|u\|_2 = \left(\|u_1\|_2^2 + \|u_2\|_2^2 + \|u_3\|_2^2\right)^{\frac{1}{2}} &\le \|2u_1 + u_2 + u_3\|_2 \\
&\le \|u_1 + u_2\|_2 + \|u_1 + u_3\|_2 \\
&\le \delta_2 h_1(u_1 + u_2) + \delta_2 h_2(u_1 + u_3) \\
&\le 2\delta_2 g(u).
\end{aligned}$$

So $\mathcal{L}_g(1) \subset \left(\mathcal{L}_{\|.\|_2}(2\delta_2) \cap W^\perp\right)$ and $Q'_{J_1,J_2} \le Q_{J_1,J_2}$, for Q_{J_1,J_2} as given in the proposition.

At last, we need to build a uniform bound on δ_{J_1,J_2} giving rise to (36). Using Lemma 1 (ii), if $u \in W^\perp$ is decomposed according to (31), we obtain

$$\begin{aligned}
f_d(u) = f_d(u_1 + u_2 + u_3) &\le f_d(2u_1 + u_2 + u_3) + f_d(u_1) \\
&\le f_d(u_1 + u_2) + f_d(u_1 + u_3) + f_d(u_1) \\
&\le \overline{\Delta} h_1(u_1 + u_2) + \overline{\Delta} h_2(u_1 + u_3) + \delta_1 \|u_1\|_2.
\end{aligned} \tag{38}$$

Using (13), $\|u_1\|_2$ satisfies the following two inequalities

$$\begin{aligned}
\|u_1\|_2 &\le \|u_1 + u_2\|_2 \le \delta_2 h_1(u_1 + u_2), \\
\|u_1\|_2 &\le \|u_1 + u_3\|_2 \le \delta_2 h_2(u_1 + u_3).
\end{aligned}$$

Adding these inequalities, we obtain

$$\delta_1 \|u_1\|_2 \le \frac{\overline{\Delta}}{2}\left(h_1(u_1 + u_2) + h_2(u_1 + u_3)\right).$$

Using (38), we finally conclude that, for $u \in W^\perp$

$$\begin{aligned}
f_d(u) &\le \frac{3\overline{\Delta}}{2}\left(h_1(u_1 + u_2) + h_2(u_1 + u_3)\right) \\
&\le 3\overline{\Delta}\, g(u).
\end{aligned}$$

The proof is complete. $\square$

4 Sets of data yielding K-sparse solutions or sparser

For any given $K \in \{0, \dots, N\}$ and $\tau > 0$, we introduce the subset $\mathcal{I}^\tau(K)$ as it follows:

$$\mathcal{I}^\tau(K) \stackrel{\text{def}}{=} \big\{ d \in \mathbb{R}^N \;:\; \mathrm{val}(\mathcal{P}_d) \le K \big\}. \tag{39}$$

All data belonging to $\mathcal{I}^\tau(K)$ generate a solution of $(\mathcal{P}_d)$—see (2)—which involves at most K non-zero components.

Let us define

$$G_K \stackrel{\text{def}}{=} \big\{ J \subset I : \dim(\mathcal{T}_J) \le K \big\}, \tag{40}$$

and remind that $\mathcal{T}_J = \mathrm{span}\left((\psi_j)_{j\in J}\right)$ according to (22).

The next proposition states a strong and slightly surprising result.

Proposition 3 *For any $K \in \{0, \dots, N\}$, any norm $\|.\|$ and any $\tau > 0$, we have*

$$\mathcal{I}^\tau(K) = \bigcup_{J \in G_K} \mathcal{T}_J + \mathcal{L}_{\|.\|}(\tau).$$

Some sets $\mathcal{I}^\tau(K)$, as defined in (39) and explained in the last proposition, are illustrated on Figure 1.

Proof. The case $K = 0$ is trivial ($G_0 = \{\emptyset\}$) and we assume in the following that $K \ge 1$.

Let $d \in \mathcal{I}^\tau(K)$. This means there is $(\lambda_i)_{i\in I}$—a solution of $(\mathcal{P}_d)$—that satisfies $\ell_0((\lambda_i)_{i\in I}) \le K$. Hence

$$\begin{aligned} d = \sum_{i\in J} \lambda_i \psi_i + w \qquad & \text{with} \quad w \in \mathcal{L}_{\|.\|}(\tau) \\ & \text{and} \quad J = \{ i \in I : \lambda_i \neq 0 \} \quad \text{with} \quad \#J \le K. \end{aligned}$$

Consequently $\dim(\mathcal{T}_J) \le \#J \le K$, which implies that $d \in \cup_{J\in G_K} \mathcal{T}_J + \mathcal{L}_{\|.\|}(\tau)$.

Conversely, let $d \in \cup_{J\in G_K} \mathcal{T}_J + \mathcal{L}_{\|.\|}(\tau)$, then $d = v + w$ where $v \in \cup_{J\in G_K} \mathcal{T}_J$ and $w \in \mathcal{L}_{\|.\|}(\tau)$. Then:

- $\exists\, J \subset I$ such that $v \in \mathcal{T}_J$ and the latter satisfies $\dim(\mathcal{T}_J) \le K$;
- there are real numbers $(\lambda_i)_{i\in J}$ involving at most $\dim(\mathcal{T}_J)$ non-zero components (hence $\ell_0((\lambda_i)_{i\in J}) \le \dim(\mathcal{T}_J) \le K$) such that $v = \sum_{i\in J} \lambda_i \psi_i$.
- $w \in \mathcal{L}_{\|.\|}(\tau)$ means that $\|w\| \le \tau$.

It follows that $d = \sum_{i\in J} \lambda_i \psi_i + w \in \mathcal{I}^\tau(K)$. $\square$

Given $J \subset I$, remind that $\mathcal{T}_J = \mathrm{span}\left((\psi_j)_{j\in J}\right)$—see (22). Since $(\psi_i)_{i\in I}$ is a general family of vectors, there may be numerous subsets J_n, $n = 1, 2, \dots$, such that $\mathcal{T}_{J_n} = \mathcal{T}_{J_m}$ and $J_n \neq J_m$. A non-redundant listing of all possible subspaces $\mathcal{T}_J$ when J runs over all subsets of I can be obtained with the help of the notations below.

For any $K = 0, \dots, N$, define $\mathcal{J}(K)$ by the following three properties:

$$\begin{cases} (a) & \mathcal{J}(K) \subset \big\{J \subset I : \dim(\mathcal{T}_J) = K\big\}; \\ (b) & J_1, J_2 \in \mathcal{J}(K) \ \text{ and } \ J_1 \neq J_2 \quad \Longrightarrow \quad \mathcal{T}_{J_1} \neq \mathcal{T}_{J_2}; \\ (c) & \mathcal{J}(K) \text{ is maximal:} \\ & \text{if } \ J_1 \subset I \ \text{ yields } \ \dim(\mathcal{T}_{J_1}) = K \ \text{ then } \ \exists J \in \mathcal{J}(K) \ \text{ such that } \ \mathcal{T}_J = \mathcal{T}_{J_1}. \end{cases} \tag{41}$$

Notice that in particular, $\mathcal{J}(0) = \{\emptyset\}$ and $\#\mathcal{J}(N) = 1$. One can observe that G_K, as defined in (40), satisfies

$$G_K \supset \bigcup_{k=0}^{K} \mathcal{J}(k)$$

and

$$\{\mathcal{T}_J : J \in G_K\} = \{\mathcal{T}_J : J \in \mathcal{J}(k) \text{ for } k \in \{0, \dots, K\}\}. \tag{42}$$

Using these notations, we can give a more convenient formulation of Proposition 3.

Theorem 1 *For any $K \in \{0, \dots, N\}$, any norm $\|.\|$ and any $\tau > 0$, we have*

$$\mathcal{I}^{\tau}(K) = \bigcup_{J \in \mathcal{J}(K)} \mathcal{T}_J^{\tau},$$

where we remind that for any $J \subset I$ and $\tau > 0$, $\mathcal{T}_J^{\tau}$ is defined by (23), and $\mathcal{J}(K)$ is defined by (41).

As a consequence, $\mathcal{I}^{\tau}(K)$ is closed and measurable.

Proof. The case $J = \emptyset$ (and $K = 0$) is trivial because of the convention $\text{span}(\emptyset) = \{0\}$ and $\mathcal{J}(0) = \{\emptyset\}$.

Let us first prove that $\mathcal{I}^{\tau}(K) = \bigcup_{J \in G_K} \mathcal{T}_J^{\tau}$. Using Proposition 3,

$$\mathcal{I}^{\tau}(K) = \left(\bigcup_{J \in G_K} \mathcal{T}_J\right) + \mathcal{L}_{\|.\|}(\tau) = \bigcup_{J \in G_K} \big(\mathcal{T}_J + \mathcal{L}_{\|.\|}(\tau)\big).$$

The last equality above is a trivial observation. Using Lemma 2, this summarizes us

$$\mathcal{I}^{\tau}(K) = \bigcup_{J \in G_K} \mathcal{T}_J^{\tau}.$$

Using (42), we deduce that

$$\{\mathcal{T}_J^{\tau} : J \in G_K\} = \big\{\mathcal{T}_J^{\tau} : J \in \mathcal{J}(k) \text{ for } k \in \{0, \dots, K\}\big\},$$

and therefore,

$$\mathcal{I}^{\tau}(K) = \bigcup_{k=0}^{K} \bigcup_{J \in \mathcal{J}(k)} \mathcal{T}_J^{\tau}.$$

Moreover, for any $k < K$ and $J \in \mathcal{J}(k)$, we can find $J_1 \in \mathcal{J}(K)$ such that $\mathcal{T}_J \subset \mathcal{T}_{J_1}$. Using Lemma 2, we find that $\mathcal{T}_J^{\tau} \subset \mathcal{T}_{J_1}^{\tau}$. Consequently,

$$\mathcal{I}^{\tau}(K) = \bigcup_{J \in \mathcal{J}(K)} \mathcal{T}_J^{\tau}.$$

This completes the proof of the first statement.

By Proposition 1, $\mathcal{I}^{\tau}(K)$ is a finite union of closed measurable sets, hence it is closed and measurable as well. $\square$

For any $K = 0, \ldots, N$, define the constants $\hat{\delta}_K$ and $\overline{\boldsymbol{C}}_K$ as it follows:

$$\hat{\delta}_K \overset{\text{def}}{=} \max_{J \in \mathcal{J}(K)} \delta_J, \tag{43}$$

$$\overline{\boldsymbol{C}}_K \overset{\text{def}}{=} \sum_{J \in \mathcal{J}(K)} C_J, \tag{44}$$

where $\delta_J \in [0, \overline{\Delta}]$ and C_J are the constants exhibited in Corollary 1, assertion (ii). Clearly,

$$0 \leq \hat{\delta}_K \leq \overline{\Delta}. \tag{45}$$

In particular,

$$\overline{\boldsymbol{C}}_0 = \mathbb{L}^N\left(\mathcal{L}_{\|.\|}(1)\right) \text{ and } \overline{\boldsymbol{C}}_N = \mathbb{L}^N\left(\mathcal{L}_{f_d}(1)\right). \tag{46}$$

With $\mathcal{J}(K)$, let us associate the family of subsets :

$$\mathcal{H}(K,k) \overset{\text{def}}{=} \Big\{(J_1, J_2) \in \mathcal{J}(K)^2 \;\text{ such that }\; \dim(\mathcal{T}_{J_1} \cap \mathcal{T}_{J_2}) = k\Big\}, \tag{47}$$

where $K = 1, 2, \ldots, N$ and $k = 0, 1, \ldots, K-1$.

Notice that $\mathcal{H}(K,k)$ may be empty for some k. Consider $(J_1, J_2) \in \mathcal{J}(K)^2$ such that

$$\begin{array}{rcccccccc} \mathcal{T}_{J_1} + \mathcal{T}_{J_2} = & (\mathcal{T}_{J_1} \cap \mathcal{T}_{J_2}) & \oplus & (\mathcal{T}_{J_1} \cap \mathcal{T}_{J_2}{}^{\perp}) & \oplus & (\mathcal{T}_{J_2} \cap \mathcal{T}_{J_1}{}^{\perp}) & \subset & \mathbb{R}^N \\ \dim(\mathcal{T}_{J_1} + \mathcal{T}_{J_2}) = & k & + & (K-k) & + & (K-k) & \leq & N \end{array}$$

and $k \geq 2K - N$. We see that

$$\mathcal{H}(K,k) \neq \emptyset \quad \Rightarrow \quad k \geq 2K - N.$$

Conversely,

$$k < k_K \overset{\text{def}}{=} \max\{0,\ 2K - N\} \quad \Rightarrow \quad \mathcal{H}(K,k) = \emptyset. \tag{48}$$

Notice that $\mathcal{H}(N,k) = \emptyset$, for all $k = 0, \ldots, N-1$ and that for any $K = 1, \ldots, N-1$, we have $0 \leq k_K \leq K-1$.

For $K \in \{1, \ldots, N-1\}$ and $k \in \{k_K, \ldots, K-1\}$ let us define

$$\hat{\delta}'_{K,k} \overset{\text{def}}{=} \max\left\{0, \max_{(J_1,J_2) \in \mathcal{H}(K,k)} \delta_{J_1,J_2}\right\}, \tag{49}$$

$$\overline{\boldsymbol{Q}}_{K,k} \overset{\text{def}}{=} \sum_{(J_1,J_2) \in \mathcal{H}(K,k)} Q_{J_1,J_2} \tag{50}$$

where Q_{J_1,J_2} and $\delta_{J_1,J_2} \in [0, 3\overline{\Delta}]$ are as in Proposition 2. It is clear that if $\mathcal{H}(K,k) = \emptyset$ then we find $\overline{\boldsymbol{Q}}_{K,k} = 0$ and $\hat{\delta}'_{K,k} = 0$. It follows that for any $K = 1, \ldots, N-1$ and any $k = k_K, \ldots, K-1$

$$0 \leq \hat{\delta}'_{K,k} \leq 3\overline{\Delta}. \tag{51}$$

Last, define

$$\Delta_K \stackrel{\text{def}}{=} \begin{cases} \hat{\delta}_0 & \text{, if } K = 0 \\ \max\left\{\Delta_{K-1}, \hat{\delta}_K, \max\limits_{k_K \le k \le K-1} \hat{\delta}'_{K,k}\right\} & \text{, if } 0 < K < N \\ \max\left\{\Delta_{N-1}, \hat{\delta}_N\right\} & \text{, if } K = N \end{cases} \tag{52}$$

Using (45) and (51),

$$0 \le \Delta_K \le 3\overline{\Delta}. \tag{53}$$

All these constants, introduced between (43) and (52), depend only on the family $(\psi_i)_{i\in I}$, the norms $\|.\|$ and f_d, K and k. Their upper bounds using $\overline{\Delta}$ only depend on $\|.\|$ and f_d. They are involved in the theorem below which provides a critical result in this work.

Theorem 2 *Let $K \in \{0, \ldots, N\}$, the norms $\|.\|$ and f_d, and $(\psi_i)_{i\in I}$, be any. Let $\tau > 0$ and $\theta \ge \tau\Delta_K$ where Δ_K is defined in (52). The Lebesgue measure in $\mathbb{R}^N$ of the set $\mathcal{I}^\tau(K)$ defined by (39) satisfies*

$$\overline{\boldsymbol{C}}_K \tau^{N-K}(\theta - \hat{\delta}_K\tau)^K - \theta^N\ \varepsilon_0(K,\tau,\theta)\ \le\ \mathbb{L}^N\left(\mathcal{I}^\tau(K) \cap \mathcal{L}_{f_d}(\theta)\right)\ \le\ \overline{\boldsymbol{C}}_K \tau^{N-K}(\theta + \hat{\delta}_K\tau)^K, \tag{54}$$

where

$$\varepsilon_0(K,\tau,\theta) = \begin{cases} 0 & \text{if } K = 0 \ \text{ or } K = N \\ \displaystyle\sum_{k=k_K}^{K-1} \overline{\boldsymbol{Q}}_{K,k}\left(\frac{\tau}{\theta}\right)^{N-k}\left(1 + \hat{\delta}'_{K,k}\frac{\tau}{\theta}\right)^k & \text{if } 0 < K < N \end{cases} \tag{55}$$

for $\overline{\boldsymbol{C}}_K$, k_K, $\overline{\boldsymbol{Q}}_{K,k}$, $\hat{\delta}_k$ and $\hat{\delta}'_{K,k}$ defined by (44), (48), (50), (43) *and* (49) *respectively. Moreover,* (45), (51) *and* (53) *provide bounds on $\hat{\delta}_K$, $\hat{\delta}'_{K,k}$ and Δ_K, respectively, which depend only on $\|.\|$ and f_d, via $\overline{\Delta}$ (see Lemma 1 (iii)).*

Remark 6 *We posit the assumptions of Theorem 2. Then asymptotically*

$$\mathbb{L}^N\left(\mathcal{I}^\tau(K) \cap \mathcal{L}_{f_d}(\theta)\right) = \overline{\boldsymbol{C}}_K\ \theta^N\left(\frac{\tau}{\theta}\right)^{N-K} + \theta^N\ o\left(\left(\frac{\tau}{\theta}\right)^{N-K}\right) \quad as \quad \frac{\tau}{\theta} \to 0.$$

Proof. Using Theorem 1, it is straightforward that

$$\mathcal{I}^\tau(K) \cap \mathcal{L}_{f_d}(\theta)\ =\ \bigcup_{J\in\mathcal{J}(K)}\left(\mathcal{T}_J^\tau \cap \mathcal{L}_{f_d}(\theta)\right) \tag{56}$$

and that

$$\mathbb{L}^N\left(\mathcal{I}^\tau(K) \cap \mathcal{L}_{f_d}(\theta)\right) = \mathbb{L}^N\Big(\bigcup_{J\in\mathcal{J}(K)}\left(\mathcal{T}_J^\tau \cap \mathcal{L}_{f_d}(\theta)\right)\Big). \tag{57}$$

When $K = 0$ or $K = N$, we have $\#\mathcal{J}(K) = 1$. Then, (54) is a straightforward consequence of (57) and Proposition 1 (the latter can be applied thanks to the assumption $\theta > \tau\Delta_K$ and (52)).

The rest of the proof is to find relevant bounds for the right-hand side of (57) under the assumption that $0 < K < N$.

Upper bound. By the definition of a measure, and then using Corollary 1, it is found that

$$\begin{aligned}
\mathbb{L}^N\left(\mathcal{I}^\tau(K)\cap\mathcal{L}_{f_d}(\theta)\right) &\le \sum_{J\in\mathcal{J}(K)}\mathbb{L}^N\left(\mathcal{T}_J^\tau\cap\mathcal{L}_{f_d}(\theta)\right) \\
&\le \tau^{N-K}\sum_{J\in\mathcal{J}(K)} C_J(\theta+\delta_J\tau)^K \\
&\le \tau^{N-K}(\theta+\hat{\delta}_K\tau)^K\sum_{J\in\mathcal{J}(K)} C_J \\
&= \overline{\boldsymbol{C}}_K\,\tau^{N-K}(\theta+\hat{\delta}_K\tau)^K,
\end{aligned}\tag{58}$$

where the constants $\hat{\delta}_K$ and $\overline{\boldsymbol{C}}_K$ are defined in (43) and (44), respectively.

Lower bound. First we represent the right-hand side of (56) as a union of disjoint subsets. Since $\mathcal{J}(K)$ is finite, let us enumerate its elements as

$$\mathcal{J}(K)=\{J_1,\dots,J_M\}\ \text{ where }\ M=\#\big(\mathcal{J}(K)\big).$$

To simplify the expressions that follow, for any J we denote

$$B_J=\mathcal{T}_J^\tau\cap\mathcal{L}_{f_d}(\theta).\tag{59}$$

Then

$$\bigcup_{J\in\mathcal{J}(K)}\Big(\mathcal{T}_J^\tau\cap\mathcal{L}_{f_d}(\theta)\Big)=\bigcup_{i=1}^M B_{J_i}.$$

Consider the following decomposition:

$$\begin{aligned}
\bigcup_{i=1}^M B_{J_i} &= \Big(B_{J_1}\Big)\cup\Big(B_{J_2}\setminus(B_{J_1}\cap B_{J_2})\Big)\cup\ldots\cup\Big(B_{J_M}\setminus\big(\cup_{j=1}^{M-1}(B_{J_j}\cap B_{J_M})\big)\Big) \\
&= \Big(B_{J_1}\Big)\cup\bigcup_{i=2}^M\left(B_{J_i}\setminus\Big(\bigcup_{j=1}^{i-1}(B_{J_j}\cap B_{J_i})\Big)\right).
\end{aligned}$$

Since the last row is a union of disjoint sets, we have

$$\mathbb{L}^N\big(\bigcup_{i=1}^M B_{J_i}\big)=\mathbb{L}^N\left(B_{J_1}\right)+\sum_{i=2}^M\mathbb{L}^N\left((B_{J_i}\setminus\big(\cup_{j=1}^{i-1}(B_{J_j}\cap B_{J_i})\big)\right).$$

Noticing that $\Big(\bigcup_{j=1}^{i-1}(B_{J_j}\cap B_{J_i})\Big)\subset B_{J_i}$ entails that

$$\mathbb{L}^N\left(B_{J_i}\setminus\big(\cup_{j=1}^{i-1}(B_{J_j}\cap B_{J_i})\big)\right)=\mathbb{L}^N\left(B_{J_i}\right)-\mathbb{L}^N\left(\cup_{j=1}^{i-1}(B_{J_j}\cap B_{J_i})\right),\quad\forall i=2,\dots,M.$$

Hence

$$\mathbb{L}^N\big(\bigcup_{i=1}^M B_{J_i}\big)=\sum_{i=1}^M\mathbb{L}^N\left(B_{J_i}\right)-\sum_{i=2}^M\mathbb{L}^N\big(\bigcup_{j=1}^{i-1}(B_{J_j}\cap B_{J_i})\big).\tag{60}$$

Using successively (59), assertion (ii) of Corollary 1, (43), (44) and $\theta\ge\tau\Delta_K$ shows that

$$\begin{aligned}
\sum_{i=1}^M\mathbb{L}^N\left(B_{J_i}\right) &= \sum_{J\in\mathcal{J}(K)}\mathbb{L}^N\left(\mathcal{T}_J^\tau\cap\mathcal{L}_{f_d}(\theta)\right) \\
&\ge \sum_{J\in\mathcal{J}(K)} C_J\tau^{N-K}(\theta-\delta_J\tau)^K \\
&\ge \overline{\boldsymbol{C}}_K\tau^{N-K}(\theta-\hat{\delta}_K\tau)^K,
\end{aligned}\tag{61}$$

where the constants $\hat{\delta}_K$ and $\overline{\boldsymbol{C}}_K$ are given in (43) and (44), respectively.

Using the original notation (59), each term, for $i = 2, ..., M$, in the last sum in (60) satisfies

$$\mathbb{L}^N \big(\bigcup_{j=1}^{i-1} (B_{J_j} \cap B_{J_i}) \big) \leq \sum_{j=1}^{i-1} \mathbb{L}^N \left(B_{J_j} \cap B_{J_i} \right) = \sum_{j=1}^{i-1} \mathbb{L}^N \left(\mathcal{L}_{f_d}(\theta) \cap \mathcal{T}_{J_j}^\tau \cap \mathcal{T}_{J_i}^\tau \right). \tag{62}$$

Let us remind that $\dim(\mathcal{T}_{J_i}) = K$ for every $i = 1, \ldots, M$ and that by the definition of $\mathcal{J}(K)$—see (41)—we have $\mathcal{T}_{J_j} \neq \mathcal{T}_{J_i}$ if $i \neq j$. Proposition 2 can hence be applied to each term of the last sum:

$$\begin{aligned} \mathbb{L}^N \left(\mathcal{L}_{f_d}(\theta) \cap \mathcal{T}_{J_j}^\tau \cap \mathcal{T}_{J_i}^\tau \right) \quad &\leq \quad Q_{J_i,J_j} \tau^{N-k_{i,j}} (\theta + \delta_{J_i,J_j} \tau)^{k_{i,j}} \\ \text{where} \qquad & k_{i,j} = \dim \left(\mathcal{T}_{J_j} \cap \mathcal{T}_{J_i} \right). \end{aligned}$$

Then (62) leads to

$$\mathbb{L}^N \big(\bigcup_{j=1}^{i-1} (B_{J_j} \cap B_{J_i}) \big) \leq \sum_{j=1}^{i-1} Q_{J_j,J_i} \tau^{N-k_{i,j}} (\theta + \delta_{J_j,J_i} \tau)^{k_{i,j}}.$$

By rearranging the last sum in (60) and taking into account (48), we obtain

$$\sum_{i=2}^{M} \mathbb{L}^N \big(\bigcup_{j=1}^{i-1} (B_{J_j} \cap B_{J_i}) \big) \leq \sum_{k=k_K}^{K-1} \overline{\boldsymbol{Q}}_{K,k} \tau^{N-k} (\theta + \hat{\delta}'_{K,k} \tau)^k, \tag{63}$$

where $\hat{\delta}'_{K,k}$ and $\overline{\boldsymbol{Q}}_{K,k}$ are given in (49) and (50), respectively.

Combining (57) along with the original notations (59) and then (60), (61) and (63) yields

$$\begin{aligned} \mathbb{L}^N \left(\mathcal{I}^\tau(K) \cap \mathcal{L}_{f_d}(\theta) \right) \quad &= \quad \mathbb{L}^N \big(\bigcup_{i=1}^{M} B_{J_i} \big) \\ &\geq \quad \overline{\boldsymbol{C}}_K \tau^{N-K} (\theta - \hat{\delta}_K \tau)^K - \varepsilon_0(K, \tau, \theta), \end{aligned}$$

where $\varepsilon_0(.)$ is as in the proposition. This finishes the proof. $\square$

Remark 7 *In the proof of this theorem we could notice (see* (60)*,* (62) *and* (58)*) that*

$$\begin{aligned} \sum_{J \in \mathcal{J}(K)} \mathbb{L}^N \left(\mathcal{T}_J^\tau \cap \mathcal{L}_{f_d}(\theta) \right) \quad &- \quad \sum_{k=k_K}^{K-1} \sum_{(J_1,J_2) \in \mathcal{H}(K,k)} \mathbb{L}^N \left(\mathcal{L}_{f_d}(\theta) \cap \mathcal{T}_{J_1}^\tau \cap \mathcal{T}_{J_2}^\tau \right) \\ &\leq \quad \mathbb{L}^N \left(\mathcal{I}^\tau(K) \cap \mathcal{L}_{f_d}(\theta) \right) \\ &\leq \quad \sum_{J \in \mathcal{J}(K)} \mathbb{L}^N \left(\mathcal{T}_J^\tau \cap \mathcal{L}_{f_d}(\theta) \right). \end{aligned} \tag{64}$$

These are the main approximations of $\mathbb{L}^N \left(\mathcal{I}^\tau(K) \cap \mathcal{L}_{f_d}(\theta) \right)$ in the proof of the theorem. The precision of the bounds given in the theorem could be more accurate by improving the above inequalities. The loss of accuracy has however the same order of magnitude as the precision in the calculus of $\mathbb{L}^N \left(\mathcal{T}_J^\tau \cap \mathcal{L}_{f_d}(\theta) \right)$.

The constants Δ_K, $\hat{\delta}_K$ and $\hat{\delta}'_{K,k}$ depend on $(\psi_i)_{i \in I}$ and K. Using the uniform bound $\overline{\Delta}$ exhibited in Lemma 1 (ii) in place of $\hat{\delta}_K$ and $\hat{\delta}'_{K,k}$ leads to a more general but less precise result.

Corollary 2 *Let $K \in \{0,\dots,N\}$, the norms $\|.\|$ and f_d, and $(\psi_i)_{i\in I}$, be any. Let $\tau > 0$ and $\theta \geq 3\tau\overline{\Delta}$ where $\overline{\Delta}$ is derived in Lemma 1 (ii) and depends only on f_d and $\|.\|$. The set $\mathcal{I}^\tau(K)$ defined by (39) satisfies*

$$\overline{\boldsymbol{C}}_K\tau^{N-K}(\theta-\overline{\Delta}\tau)^K - \theta^N\ \varepsilon_0^u(K,\tau,\theta) \leq \mathbb{L}^N\left(\mathcal{I}^\tau(K)\cap\mathcal{L}_{f_d}(\theta)\right) \leq \overline{\boldsymbol{C}}_K\tau^{N-K}(\theta+\overline{\Delta}\tau)^K, \tag{65}$$

where

$$\varepsilon_0^u(K,\tau,\theta) = \begin{cases} 0 & \text{, if } K=0 \text{ or } K=N \\ \displaystyle\sum_{k=k_K}^{K-1} \overline{\boldsymbol{Q}}_{K,k}\left(\frac{\tau}{\theta}\right)^{N-k}\left(1+3\overline{\Delta}\,\frac{\tau}{\theta}\right)^k & \text{, if } 0<K<N. \end{cases}$$

Moreover, for $K=1,\dots,N-1$ and $k=k_K,\dots,K-1$, we have

$$\overline{\boldsymbol{Q}}_{K,k} \leq \#\mathcal{J}(K)(\#\mathcal{J}(K)-1)\alpha(N-k)\alpha(k)(2\delta_2)^{N-k}\delta_3^k \tag{66}$$

where

$$\#\mathcal{J}(K) \leq \frac{\#I!}{K!(\#I-K)!}, \tag{67}$$

$\alpha(n)$ is the volume of unit ball for the euclidean norm in $\mathbb{R}^n$ (see equation (78) *for details), δ_2 is defined in Lemma 1 (see equation* (10)*) and δ_3 is such that*

$$\|w\|_2 \leq \delta_3 f_d(w)\ , \qquad \forall w\in\mathbb{R}^N.$$

Proof. Equation (65) is obtained by inserting in (54) in Theorem 2 the uniform bounds on $\hat{\delta}_K$, $\hat{\delta}'_{K,k}$ and Δ_K given in (45), (51) and (53), respectively.

The upper bound for $\overline{\boldsymbol{Q}}_{K,k}$ is obtained as follows. Using (50) and (27), we obtain

$$\overline{\boldsymbol{Q}}_{K,k} = \sum_{(J_1,J_2)\in\mathcal{H}(K,k)} \mathbb{L}^{N-k}\left((\mathcal{T}_{J_1}\cap\mathcal{T}_{J_2})^\perp\cap\mathcal{L}_{\|.\|_2}(2\delta_2)\right)\mathbb{L}^k\left(\mathcal{T}_{J_1}\cap\mathcal{T}_{J_2}\cap\mathcal{L}_{f_d}(1)\right).$$

Moreover,

$$\mathbb{L}^{N-k}\left((\mathcal{T}_{J_1}\cap\mathcal{T}_{J_2})^\perp\cap\mathcal{L}_{\|.\|_2}(2\delta_2)\right) = \alpha(N-k)(2\delta_2)^{N-k},$$
$$\mathbb{L}^k\left(\mathcal{T}_{J_1}\cap\mathcal{T}_{J_2}\cap\mathcal{L}_{f_d}(1)\right) \leq \mathbb{L}^k\left(\mathcal{T}_{J_1}\cap\mathcal{T}_{J_2}\cap\mathcal{L}_{\|.\|_2}(\delta_3)\right) = \alpha(k)(\delta_3)^k,$$

and we obviously have

$$\#\mathcal{H}(K,k) \leq \#\mathcal{J}(K)(\#\mathcal{J}(K)-1).$$

□

The above corollary shows that the "quality" of the asymptotic as $\frac{\tau}{\theta}\to 0$ depends on $\|.\|$, f_d and on the dictionary through the terms $\overline{\boldsymbol{Q}}_{K,k}$. The latter terms are bounded from above using (66) and (67) and they are clearly overestimated. Even though the bound we provide are very pessimistic, they depend only on $\|.\|$, f_d and $\#I$ and can be computed.

Remark 8 *Let us emphasize that "uniform" bounds in the spirit of Corollary 2 can be derived from Proposition 4, and Theorems 3, 4 and 5. We leave this task to interested readers that need to compute easily the relevant bounds.*

5 Sets of data yielding K-sparse solutions

For any $K \in \{0,\dots,N\}$ and $\tau > 0$, we denote

$$\mathcal{D}^\tau(K) \stackrel{\text{def}}{=} \left\{d \in \mathbb{R}^N : \mathrm{val}(\mathcal{P}_d) = K\right\}. \tag{68}$$

¿From the definition of $\mathcal{I}^\tau(K)$ in (39), it is straightforward that

$$\mathcal{D}^\tau(K) = \mathcal{I}^\tau(K) \setminus \mathcal{I}^\tau(K-1), \qquad \forall K \in \{0,\dots,N\}, \tag{69}$$

where we extend the definition of $\mathcal{I}^\tau(K)$ with

$$\mathcal{I}^\tau(-1) = \emptyset.$$

Being the difference of two measurable closed sets, $\mathcal{D}^\tau(K)$ is clearly measurable. Noticing also that

$$\mathcal{I}^\tau(K-1) \subset \mathcal{I}^\tau(K) \tag{70}$$

we get

$$\mathbb{L}^N\left(\mathcal{D}^\tau(K) \cap \mathcal{L}_{f_d}(\theta)\right) = \mathbb{L}^N\left(\mathcal{I}^\tau(K) \cap \mathcal{L}_{f_d}(\theta)\right) - \mathbb{L}^N\left(\mathcal{I}^\tau(K-1) \cap \mathcal{L}_{f_d}(\theta)\right). \tag{71}$$

Combining these observations with Theorem 2 yields an important statement which is given below.

Theorem 3 *Let $K \in \{0,\dots,N\}$, the norms $\|.\|$ and f_d, and $(\psi_i)_{i\in I}$, be any. Let $\theta > 0$ and $\theta \geq \tau\max(\Delta_K, \Delta_{K-1})$ where Δ_k is defined in (52), for $k \in \{K-1, K\}$. The Lebesgue measure in $\mathbb{R}^N$ of the set $\mathcal{D}^\tau(K)$ defined in (68) satisfies*

$$\begin{aligned}\overline{\boldsymbol{C}}_K\ \tau^{N-K}(\theta - \hat{\delta}_K\tau)^K - \theta^N\varepsilon_0'(K,\tau,\theta) \quad &\leq \quad \mathbb{L}^N\left(\mathcal{D}^\tau(K) \cap \mathcal{L}_{f_d}(\theta)\right) &(72)\\ &\leq \quad \overline{\boldsymbol{C}}_K\ \tau^{N-K}(\theta + \hat{\delta}_K\tau)^K + \theta^N\varepsilon_1(K,\tau,\theta), &(73)\end{aligned}$$

with

$$\begin{aligned}\varepsilon_0'(K,\tau,\theta) \quad &= \quad \varepsilon_0(K,\tau,\theta) + \overline{\boldsymbol{C}}_{K-1}\left(\frac{\tau}{\theta}\right)^{N-(K-1)}\left(1+\hat{\delta}_{K-1}\frac{\tau}{\theta}\right)^{K-1},\\ \varepsilon_1(K,\tau,\theta) \quad &= \quad \varepsilon_0(K-1,\tau,\theta) - \overline{\boldsymbol{C}}_{K-1}\left(\frac{\tau}{\theta}\right)^{N-(K-1)}\left(1-\hat{\delta}_{K-1}\frac{\tau}{\theta}\right)^{K-1},\end{aligned}$$

where $\overline{\boldsymbol{C}}_k$ for $k \in \{K-1, K\}$ are defined by (44), along with the extension $\overline{\boldsymbol{C}}_{-1} = 0$, whereas ε_0 is as in Theorem 2 with the extension $\varepsilon_0(-1,\tau,\theta) \equiv 0$.

Proof. By (71), we have

$$\begin{aligned}\mathbb{L}^N\left(\mathcal{D}^\tau(K) \cap \mathcal{L}_{f_d}(\theta)\right) \quad &\leq \quad \text{Upper bound}\Big(\mathbb{L}^N\left(\mathcal{I}^\tau(K) \cap \mathcal{L}_{f_d}(\theta)\right)\Big) - \text{Lower bound}\Big(\mathbb{L}^N\left(\mathcal{I}^\tau(K-1) \cap \mathcal{L}_{f_d}(\theta)\right)\Big)\\ \mathbb{L}^N\left(\mathcal{D}^\tau(K) \cap \mathcal{L}_{f_d}(\theta)\right) \quad &\geq \quad \text{Lower bound}\Big(\mathbb{L}^N\left(\mathcal{I}^\tau(K) \cap \mathcal{L}_{f_d}(\theta)\right)\Big) - \text{Upper bound}\Big(\mathbb{L}^N\left(\mathcal{I}^\tau(K-1) \cap \mathcal{L}_{f_d}(\theta)\right)\Big)\end{aligned}$$

where the relevant upper and lower bounds were derived in Theorem 2. Since $\mathbb{L}^N\left(\mathcal{I}^\tau(K-1) \cap \mathcal{L}_{f_d}(\theta)\right)$ is negligible compared to $\mathbb{L}^N\left(\mathcal{I}^\tau(K) \cap \mathcal{L}_{f_d}(\theta)\right)$, the bounds corresponding to this term are introduced in the error functions $\varepsilon_0'(K,\tau,\theta)$ and $\varepsilon_1(K,\tau,\theta)$. $\square$

Remark 9 *Let us emphasize that Remark 6 is valid if we write* $\mathcal{D}^\tau(K)$ *in place of* $\mathcal{I}^\tau(K)$*. This gives the asymptotic of the* $\mathbb{L}^N\left(\mathcal{D}^\tau(K)\cap\mathcal{L}_{f_d}(\theta)\right)$ *as* $\frac{\tau}{\theta}$ *goes to* 0*. This observation may seem surprising. It only means that as far as* $\frac{\tau}{\theta}$ *decreases, the chance to get a solution with sparsity* strictly *smaller than* K *is very small when compared to the chance of getting a sparsity* K*.*

Remark 10 *In Section 4, we adapted Theorem 2 to get Corollary 2. In the latter, the gap between the lower and upper bounds only depends on* $\|.\|$*,* f_d *and* $\overline{\boldsymbol{Q}}_{K,K-1}$ *the latter depending on the dictionary in a controllable way. A similar adaptation of Theorem 3 is easy.*

6 Statistical meaning of the results

In this section we give a statistical interpretation of our main results, namely Theorem 2 and Theorem 3.

Proposition 4 *Let* f_d *and* $\|.\|$ *be any two norms and* $(\psi_i)_{i\in I}$ *be a dictionary in* $\mathbb{R}^N$*. For any* $K\in\{0,\ldots,N\}$*, let* $\tau>0$ *and* θ *be such that* $\theta\geq\tau\Delta_K$ *where* Δ_k *is defined in (52). Consider a random variable* d *with uniform distribution on* $\mathcal{L}_{f_d}(\theta)$*. Then*

$$\begin{aligned}\frac{\overline{\boldsymbol{C}}_K}{\mathbb{L}^N\left(\mathcal{L}_{f_d}(1)\right)}\left(\frac{\tau}{\theta}\right)^{N-K}\left(1-\hat{\delta}_K\frac{\tau}{\theta}\right)^K-\frac{\varepsilon_0(K,\tau,\theta)}{\mathbb{L}^N\left(\mathcal{L}_{f_d}(1)\right)} &\leq \mathbb{P}\left(\mathrm{val}(\mathcal{P}_d)\leq K\right)\\ &\leq \frac{\overline{\boldsymbol{C}}_K}{\mathbb{L}^N\left(\mathcal{L}_{f_d}(1)\right)}\left(\frac{\tau}{\theta}\right)^{N-K}\left(1+\hat{\delta}_K\frac{\tau}{\theta}\right)^K,\end{aligned}$$

where $\varepsilon_0(K,\tau,\theta)$ *is given in Theorem 2, equation* (55)*. Moreover we have the following asymptotical result:*

$$\mathbb{P}\left(\mathrm{val}(\mathcal{P}_d)\leq K\right)=\frac{\overline{\boldsymbol{C}}_K}{\mathbb{L}^N\left(\mathcal{L}_{f_d}(1)\right)}\left(\frac{\tau}{\theta}\right)^{N-K}+o\left(\left(\frac{\tau}{\theta}\right)^{N-K}\right)\quad as\ \frac{\tau}{\theta}\to 0.$$

Proof. Consider the set $\mathcal{I}^\tau(K)$ defined by (39). We have

$$\mathbb{P}\left(\mathrm{val}(\mathcal{P}_d)\leq K\right)=\mathbb{P}\left(d\in\mathcal{I}^\tau(K)\cap\mathcal{L}_{f_d}(\theta)\right)=\frac{\mathbb{L}^N\left(\mathcal{I}^\tau(K)\cap\mathcal{L}_{f_d}(\theta)\right)}{\mathbb{L}^N\left(\mathcal{L}_{f_d}(\theta)\right)},$$

since d is uniformly distributed on $\mathcal{L}_{f_d}(\theta)$. The inequality result follow from Theorem 2, equation (54) and uses the observation that $\mathbb{L}^N\left(\mathcal{L}_{f_d}(\theta)\right)=\theta^N\,\mathbb{L}^N\left(\mathcal{L}_{f_d}(1)\right)$.

The asymptotical result is a direct consequence of Remark 6. □

Remark 11 *Notice that, as already noticed in* (46)*,* $\overline{\boldsymbol{C}}_N=\mathbb{L}^N\left(\mathcal{L}_{f_d}(1)\right)$ *and the asymptotic in Proposition 4 reads for* $K=N$

$$\mathbb{P}\left(\mathrm{val}(\mathcal{P}_d)\leq N\right)=1+o(1)\qquad as\ \frac{\tau}{\theta}\to 0.$$

In fact a better estimate is easy to obtain in this particular case. We know indeed that for all $d\in\mathbb{R}^N$*, any solution of* $\mathcal{P}_d$ *involves an independent system of elements of* $(\psi_i)_{i\in I}$*. (A sparser decomposition would otherwise exist.) Therefore we know that for all* $d\in\mathbb{R}^N$*,* $\mathrm{val}(\mathcal{P}_d)\leq N$*. This yields*

$$\mathbb{P}\left(\mathrm{val}(\mathcal{P}_d)\leq N\right)=1. \tag{74}$$

Theorem 4 *Let f_d and $\|.\|$ be any two norms and $(\psi_i)_{i\in I}$ be a dictionary in $\mathbb{R}^N$. For any $K \in \{0,\dots,N\}$, let $\tau > 0$ and θ be such that $\theta \geq \tau \max(\Delta_K, \Delta_{K-1})$ where Δ_k is defined in (52). Consider a random variable d with uniform distribution on $\mathcal{L}_{f_d}(\theta)$. Then we have*

$$\begin{aligned} \frac{\overline{\boldsymbol{C}}_K}{\mathbb{L}^N\left(\mathcal{L}_{f_d}(1)\right)}\left(\frac{\tau}{\theta}\right)^{N-K}\left(1-\hat{\delta}_K\frac{\tau}{\theta}\right)^K - \varepsilon^-(K,\tau,\theta) \quad &\leq \quad \mathbb{P}\left(\mathrm{val}(\mathcal{P}_d)=K\right) \\ &\leq \quad \frac{\overline{\boldsymbol{C}}_K}{\mathbb{L}^N\left(\mathcal{L}_{f_d}(1)\right)}\left(\frac{\tau}{\theta}\right)^{N-K}\left(1+\hat{\delta}_K\frac{\tau}{\theta}\right)^K + \varepsilon^+(K,\tau,\theta) \end{aligned}$$

with

$$\varepsilon^-(K,\tau,\theta) = \frac{\varepsilon_0'(K,\tau,\theta)}{\mathbb{L}^N\left(\mathcal{L}_{f_d}(1)\right)}$$

and

$$\varepsilon^+(K,\tau,\theta) = \frac{\varepsilon_1(K,\tau,\theta)}{\mathbb{L}^N\left(\mathcal{L}_{f_d}(1)\right)},$$

for ε_0' and ε_1 as defined in Theorem 3 and for $\hat{\delta}_K$ and $\overline{\boldsymbol{C}}_K$ defined in (43) and (44), respectively.

In particular, we have

$$\mathbb{P}\left(\mathrm{val}(\mathcal{P}_d)=K\right) = \frac{\overline{\boldsymbol{C}}_K}{\mathbb{L}^N\left(\mathcal{L}_{f_d}(1)\right)}\left(\frac{\tau}{\theta}\right)^{N-K} + o\left(\left(\frac{\tau}{\theta}\right)^{N-K}\right) \quad \text{as} \quad \frac{\tau}{\theta}\to 0. \tag{75}$$

Proof. Consider the set $\mathcal{D}^\tau(K)$ defined in (68). We have

$$\mathbb{P}\left(\mathrm{val}(\mathcal{P}_d)=K\right) = \mathbb{P}\left(d\in\mathcal{D}^\tau(K)\cap\mathcal{L}_{f_d}(\theta)\right) = \frac{\mathbb{L}^N\left(\mathcal{D}^\tau(K)\cap\mathcal{L}_{f_d}(\theta)\right)}{\mathbb{L}^N\left(\mathcal{L}_{f_d}(\theta)\right)},$$

since d is uniformly distributed on $\mathcal{L}_{f_d}(\theta)$. The inequality result follows from Theorem 3, equation (73), and $\mathbb{L}^N\left(\mathcal{L}_{f_d}(\theta)\right) = \mathbb{L}^N\left(\mathcal{L}_{f_d}(1)\right)\theta^N$. □

Remark 12 *¿From (75) and (46), we see that*

$$\mathbb{P}\left(\mathrm{val}(\mathcal{P}_d)=N\right) = 1 + o(1) \qquad \text{as } \frac{\tau}{\theta}\to 0.$$

For any other $K \in \{0,\dots,N-1\}$, $\mathbb{P}\left(\mathrm{val}(\mathcal{P}_d)=K\right)$ goes to 0, as $\frac{\tau}{\theta}\to 0$. Moreover, we know how rapidly they go to 0. In particular, we know that $\mathbb{P}\left(\mathrm{val}(\mathcal{P}_d)=K-1\right)$ becomes negligible when compared to $\mathbb{P}\left(\mathrm{val}(\mathcal{P}_d)=K\right)$, as $\frac{\tau}{\theta}\to 0$.

Notice that even though d is a random variable on a subset of $\mathbb{R}^N$, the value of our function $\mathrm{val}(\mathcal{P}_d)$ is an integer larger than zero. We can also compute the expectation of $\mathrm{val}(\mathcal{P}_d)$:

$$\begin{aligned}
\mathbb{E}\left(\mathrm{val}(\mathcal{P}_d)\right) &= \sum_{K=1}^{N} K\ \mathbb{P}\left(\mathrm{val}(\mathcal{P}_d) = K\right) \\
&= \sum_{K=1}^{N} K\left(\mathbb{P}\left(\mathrm{val}(\mathcal{P}_d) \leq K\right) - \mathbb{P}\left(\mathrm{val}(\mathcal{P}_d) \leq K-1\right)\right) \\
&= \sum_{K=0}^{N} K\ \mathbb{P}\left(\mathrm{val}(\mathcal{P}_d) \leq K\right) - \sum_{K=0}^{N-1} (K+1)\mathbb{P}\left(\mathrm{val}(\mathcal{P}_d) \leq K\right) \\
&= \mathbb{P}\left(\mathrm{val}(\mathcal{P}_d) \leq N\right) - \sum_{K=0}^{N-1} \mathbb{P}\left(\mathrm{val}(\mathcal{P}_d) \leq K\right) \\
&= N - \sum_{K=0}^{N-1} \mathbb{P}\left(\mathrm{val}(\mathcal{P}_d) \leq K\right)
\end{aligned}$$

where we used (70) and (74).

This yields the following Theorem.

Theorem 5 *Let f_d and $\|.\|$ be any two norms and $(\psi_i)_{i\in I}$ be a dictionary in $\mathbb{R}^N$. Let $\tau > 0$ and θ be such that $\theta \geq \tau \max_{0\leq K\leq N} \Delta_K$ where Δ_K is defined in (52). Consider a random variable d with uniform distribution on $\mathcal{L}_{f_d}(\theta)$. Then*

$$\begin{aligned}
N - \sum_{K=0}^{N-1} \frac{\overline{\boldsymbol{C}}_K}{\mathbb{L}^N\left(\mathcal{L}_{f_d}(1)\right)} \left(\frac{\tau}{\theta}\right)^{N-K} \left(1 + \hat{\delta}_K \frac{\tau}{\theta}\right)^K \quad &\leq \quad \mathbb{E}\left(\mathrm{val}(\mathcal{P}_d)\right) \\
&\leq \quad N - \sum_{K=0}^{N-1} \frac{\overline{\boldsymbol{C}}_K}{\mathbb{L}^N\left(\mathcal{L}_{f_d}(1)\right)} \left(\frac{\tau}{\theta}\right)^{N-K} \left(1 - \hat{\delta}_K \frac{\tau}{\theta}\right)^K - \frac{\varepsilon_0(K,\tau,\theta)}{\mathbb{L}^N\left(\mathcal{L}_{f_d}(1)\right)}
\end{aligned}$$

where $\varepsilon_0(K,\tau,\theta)$ is given in Theorem 2, equation (55). Moreover we have the following asymptotical result:

$$\mathbb{E}\left(\mathrm{val}(\mathcal{P}_d)\right) = N - \frac{\overline{\boldsymbol{C}}_{N-1}}{\mathbb{L}^N\left(\mathcal{L}_{f_d}(1)\right)} \frac{\tau}{\theta} + o\left(\frac{\tau}{\theta}\right) \quad \text{as} \quad \frac{\tau}{\theta} \to 0.$$

7 Illustration: Euclidean norms for $\|.\|$ and f_d

Consider the situation when both $\|.\|$ and f_d are the Euclidean norm on $\mathbb{R}^N$:

$$\|.\| = f_d = \|.\|_2 \quad \text{where} \quad \|u\|_2 = \sqrt{\langle u,u\rangle}, \quad \text{with} \quad \langle u,v\rangle = \sum_{i=1}^{N} u_i v_i. \tag{76}$$

Noticing that the Euclidean norm is rotation invariant, for any vector subspace $V \subseteq \mathbb{R}^N$ we have

$$P_{V^\perp}\left(\mathcal{L}_{\|.\|_2}(\tau)\right) = V^\perp \cap \mathcal{L}_{\|.\|_2}(\tau) = \{u \in V^\perp : \|u\|_2 \leq \tau\}. \tag{77}$$

The equivalent norm h and the constant $\overline{\Delta}$ derived in Lemma 1 are simply

$$\begin{aligned}
h(u) &= \|u\|_2, \quad \forall u \in V^\perp, \\
\overline{\Delta} &= 1.
\end{aligned}$$

The constant δ_V in assertion (ii) of Proposition 1, defined by (20), reads $\delta_V = 1$. Then the inequality condition on θ and τ is simplified to $\theta \geq \tau$.

The constant C in (19) in the same proposition depends on K (the dimension of the subspace V) and reads (see [7, p.60] for details)

$$C = \alpha(K)\alpha(N-K) \stackrel{\text{def}}{=} \mathcal{C}(K),$$

where for any integer $n > 0$ we have

$$\alpha(n) = \frac{\pi^{n/2}}{\Gamma(n/2+1)} \quad \text{for} \quad \Gamma(n) = \int_0^\infty e^{-x}x^{n-1}dx. \tag{78}$$

Here Γ is the usual Gamma function. Using that $\Gamma(n+1) = n\Gamma(n)$, it comes

$$\mathcal{C}(K) = \frac{4\pi^{\frac{N}{2}}}{K(N-K)\Gamma\left(\frac{N-K}{2}\right)\Gamma\left(\frac{K}{2}\right)} \tag{79}$$

¿From the preceding, the constants δ_J and C_J in Corollary 1 read

$$\delta_J = 1, \quad \forall J \subset I, \tag{80}$$

$$C_J = \mathcal{C}(K), \tag{81}$$

where the expression of $\mathcal{C}(K)$ is given in (79).

The norm g arising in (32) in Proposition 2 reads

$$\begin{aligned} g(u) &= \sup\{\|u_1\|_2 + \|u_2\|_2,\ \|u_1\|_2 + \|u_3\|_2\} \\ &= \|u_1\|_2 + \sup\{\|u_2\|_2,\ \|u_3\|_2\} \end{aligned}$$

where $u = u_1 + u_2 + u_3$ is decomposed according to (31). Then

$$f_d(u) = \|u\|_2 = \|u_1\|_2 + \|u_2\|_2 + \|u_3\|_2 \leq \delta_{J_1,J_2}g(u), \quad \forall u \in W^\perp \quad \text{if} \quad \delta_{J_1,J_2} = 2$$

The constants δ_{J_1,J_2} and Q_{J_1,J_2} in Proposition 2 read

$$\delta_{J_1,J_2} = 2 \tag{82}$$

$$Q_{J_1,J_2} = \mathcal{C}(k), \tag{83}$$

where $\mathcal{C}(k)$ is defined according to (79).

For any $k = 1, \ldots, N$, the constants $\hat{\delta}_k$ and $\overline{\boldsymbol{C}}_k$ in (43)-(44) read

$$\begin{aligned} \hat{\delta}_k &= 1, \\ \overline{\boldsymbol{C}}_k &= \mathcal{C}(k)\ \#\mathcal{J}(k). \end{aligned}$$

Clearly, $\#\mathcal{J}(K)$ depends on the dictionary $(\psi_i)_{i\in I}$.

The constants $\hat{\delta}'_{K,k}$ and $\overline{\boldsymbol{Q}}_{K,k}$, introduced in (49) and (50), respectively, are

$$\hat{\delta}'_{K,k} = 2, \tag{84}$$

$$\overline{\boldsymbol{Q}}_{K,k} = \mathcal{C}(k)\ \#\mathcal{H}(K,k). \tag{85}$$

Here again, $\#\mathcal{H}(K,k)$ depends on the choice of dictionary and in any case, $\#\mathcal{H}(K,k) = 0$ for $k < k_0$ (where k_0 is defined in (48)). The constant in (52) is $\Delta_K = 2$ and the inequality (53) is satisfied.

The main inequality in Theorem 2 now reads

$$\begin{aligned} \mathcal{C}(K)\#\{\mathcal{J}(K)\}\ \tau^{N-K}(\theta-\tau)^K - \varepsilon_0(K,\tau,\theta) &\leq \mathbb{L}^N\left(\mathcal{I}^\tau(K) \cap \mathcal{L}_{f_d}(\theta)\right) \\ &\leq \mathcal{C}(K)\#\{\mathcal{J}(K)\}\ \tau^{N-K}(\theta+\tau)^K, \end{aligned}$$

where $\mathcal{C}(K)$ is defined by (79) and the error term $\varepsilon_0(K,\tau,\theta)$ is

$$\varepsilon_0(K,\tau,\theta) = \frac{1}{2}\sum_{k=k_0}^{K-1} \mathcal{C}(k)\ \#\{\mathcal{H}(K,k)\}\ \tau^{N-k}(\theta+2\tau)^k.$$

In order to provide the statistical interpretation in section 6, we notice that $\mathbb{L}^N\left(\mathcal{L}_{f_d}(1)\right) = \alpha(N)$ for $\alpha(.)$ as given in (78), and hence

$$\mathbb{L}^N\left(\mathcal{L}_{f_d}(1)\right) = \frac{\pi^{N/2}}{\Gamma(N/2+1)}.$$

8 Conclusion and perspectives

In this paper, we derive lower and upper bounds for different quantities concerning the model $(\mathcal{P}_d)$. Typically, the difference between the upper and the lower bound has an order of magnitude $(\frac{\tau}{\theta})^{N-K+1}$ while the quantities which are estimated are propositional to $(\frac{\tau}{\theta})^{N-K}$. The difference between the upper and lower bounds is made of

- The terms $\theta \pm \delta_v \tau$ which come from the inclusions $B_0 \subseteq V^\tau \cap \mathcal{L}_{f_d}(\theta) \subseteq B_1$, in the proof Proposition 1. This approximation is of the order $(\frac{\tau}{\theta})^{N-K+1}$. It may be possible to reach a larger order of magnitude (e.g. $(\frac{\tau}{\theta})^{N-K+2}$) under the assumption that f_d is regular away from 0 (e.g. twice differentiable). This would permit to improve Proposition 1 and the theorems that use its conclusions.

- A term of the form $-\theta^N \varepsilon_0(K,\tau,\theta)$ could be added to the upper bound in (54). This term is not present because of the approximation made in (58). Such a term "$-\theta^N\varepsilon_0(K,\tau,\theta)$" could be obtained by computing the size of the intersection of more than two cylinder-like sets in Proposition 2 (doing so we would also avoid the approximation in (62)) and by improving this proposition by bounding $\mathbb{L}^N\left(\mathcal{T}^\tau_{J_1} \cap \mathcal{T}^\tau_{J_2} \cap \mathcal{L}_{f_d}(\theta)\right)$ from below. This is probably a straightforward adaptation of the current proof of Proposition 2.

  This improvement is possible but not necessary in this paper since (again) this approximation yields an error whose order of magnitude is $(\frac{\tau}{\theta})^{N-K+1}$. We can anyway not get a better order of magnitude unless the approximation mentioned in the previous item is not improved (i.e. more regularity is assumed for f_d).

Besides those aspects, several future developments can be envisaged:

- An important improvement would be to assume a more specialized form for the data distribution. One first step would be a distribution of the shape $\propto e^{-f_d(w)}$ which is continuous. In our opinion, one possible goal is to deal with a data distribution defined by a kernel. This is indeed one of the standard technique used in machine learning theory to approximate data distributions.

- Another way of improvement is to adapt those results to the context of infinite dimensional spaces. This adaptation might not be trivial since (for instance) there is no Lebesgue measure in those spaces.

- We are also preparing a paper where a similar analysis is performed for the Basis Pursuit Denoising (i.e. l^1 regularization) with the same asymptotic. It will clearly show what is in common and what are the differences between ℓ_0 and l^1 regularization.

- Performing a similar analysis for the Orthogonal Matching Pursuit would, of course, be a interesting and complementary result.

- In a forthcoming work, we develop the theory in the context of orthogonal bases instead of general dictionaries (frames). This simplification of the hypotheses simplifies a lot the formulas of the current paper and illustrate it.

References

[1] Julian E. Besag. On the statistical analysis of dirty pictures (with discussion). *Journal of the Royal Statistical Society B*, 48(3):259–302, 1986.

[2] Julian E. Besag. Digital image processing : Towards Bayesian image analysis. *Journal of Applied Statistics*, 16(3):395–407, 1989.

[3] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1999.

[4] R.R. Coifman and M.V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE, Transactions on Information Theory*, 38(2):713–718, March 1992.

[5] R.A. Devore. Nonlinear approximation. *Acta Numerica*, 7:51–150, 1998.

[6] David Donoho, Iain Johnstone, Jeffrey Hoch, and Alan Stern. Maximum entropy and the nearly black object. *Journal of the Royal Statistical Society B*, 54(1):41–81, 1992.

[7] L. C. Evans and R. F. Gariepy. *Measure theory and fine properties of functions.* Studies in Advanced Mathematics. CRC Press, Roca Baton, FL, 1992.

[8] Stuart Geman and Donald Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, Nov. 1984.

[9] M.J. Gormish, D. Lee, and M.W. Marcellin. Jpeg 2000: overview architecture and applications. In *proc. of ICIP 2000*, volume 2, 2000.

[10] Y.G. Leclerc. Constructing simple stable description for image partitioning. *International Journal of Computer Vision*, 3:73–102, 1989.

[11] S.Z. Li. *Markov Random Field Modeling in Computer Vision*. Springer-Verlag, New York, 1 edition, 1995.

[12] D.G. Luenberger. *Optimization by Vector Space Methods*. Wiley, J., New York, 1 edition, 1969.

[13] F. Malgouyres. Estimating the probability law of the codelength as a function of the approximation error in image compression. *Comptes Rendus de l'Acadmie des sciences, srie mathmatiques*, 344(9):607–610, 2007.

[14] F. Malgouyres. Image compression through a projection onto a polyhedral set. *Journal of Mathematical Imaging and Vision*, 27(2):193–200, Feb. 2007.

[15] F. Malgouyres. Rank related properties for basis pursuit and total variation regularization. *Signal Processing*, 87(11):2695–2707, Nov. 2007.

[16] Franois Malgouyres. Projecting onto a polytope simplifies data distributions. Technical report, University Paris 13, 2006-1, Jaunuary, 2006.

[17] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE, Transactions on Signal Processing*, 41(12):3397–3415, December 1993.

[18] Y. Pati, R. Rezaiifar, and P. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *27 th Annual Asilomar Conference on Signals, Systems, and Computers*, volume 1, pages 40–44. IEEE, 93.

[19] M.C. Robini, A. Lachal, and I.E. Magnin. A stochastic continuation approach to piecewise constant reconstruction. *IEEE Transactions on Image Processing*, 16(10):2576–2589, Oct. 2007.

[20] G.K. Wallace. The jpeg still picture compression standard. *Communications of the ACM*, 34(4):30–44, April 1991.